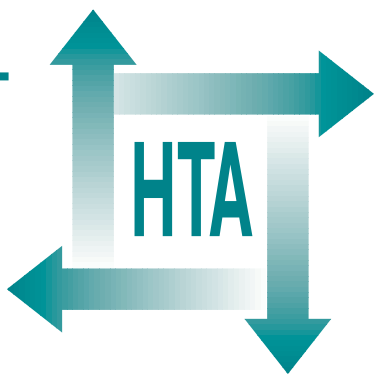
Health Technology Assessment 1998; Vol. 2: No. 16

Review

Qualitative research methods in health technology assessment: a review of the literature

E Murphy
R Dingwall
D Greatbatch
S Parker
P Watson

Health Technology Assessment
NHS R&D HTA Programme



Standing Group on Health Technology

Chair: Professor Sir Miles Irving,
Professor of Surgery, University of Manchester, Hope Hospital, Salford †

Dr Sheila Adam,
Department of Health

Professor Martin Buxton,
Professor of Economics, Brunel University †

Professor Angela Coulter,
Director, King's Fund, London

Professor Anthony Culyer,
Deputy Vice-Chancellor, University of York

Dr Peter Doyle,
Executive Director, Zeneca Ltd,
ACOST Committee on Medical Research & Health

Professor John Farndon,
Professor of Surgery, University of Bristol †

Professor Charles Florey,
Department of Epidemiology & Public Health, Ninewells Hospital & Medical School, University of Dundee †

Professor John Gabbay,
Director, Wessex Institute for Health Research & Development †

Professor Sir John Grimley Evans,
Department of Geriatric Medicine, Radcliffe Infirmary, Oxford †

Dr Tony Hope,
The Medical School, University of Oxford †

Professor Howard Glennester,
Professor of Social Science & Administration, London School of Economics & Political Science

Mr John H James,
Chief Executive, Kensington, Chelsea & Westminster Health Authority

Professor Richard Lilford,
Regional Director, R&D, West Midlands †

Professor Michael Maisey,
Professor of Radiological Sciences, UMDS, London

Dr Jeremy Metters,
Deputy Chief Medical Officer, Department of Health †

Mrs Gloria Oates,
Chief Executive, Oldham NHS Trust

Dr George Poste,
Chief Science & Technology Officer, SmithKline Beecham †

Professor Michael Rawlins,
Wolfson Unit of Clinical Pharmacology, University of Newcastle-upon-Tyne

Professor Martin Roland,
Professor of General Practice, University of Manchester

Mr Hugh Ross,
Chief Executive, The United Bristol Healthcare NHS Trust †

Professor Ian Russell,
Department of Health, Sciences & Clinical Evaluation, University of York

Professor Trevor Sheldon,
Director, NHS Centre for Reviews & Dissemination, University of York †

Professor Mike Smith,
Director, The Research School of Medicine, University of Leeds †

Dr Charles Swan,
Consultant Gastroenterologist, North Staffordshire Royal Infirmary

Dr John Tripp,
Department of Child Health, Royal Devon & Exeter Healthcare NHS Trust †

Professor Tom Walley,
Department of Pharmacological Therapeutics, University of Liverpool †

Dr Julie Woodin,
Chief Executive, Nottingham Health Authority †

† Current members

HTA Commissioning Board

Chair: Professor Charles Florey, Department of Epidemiology & Public Health,
Ninewells Hospital & Medical School, University of Dundee †

Professor Ian Russell,
Department of Health, Sciences & Clinical Evaluation, University of York *

Dr Doug Altman,
Director of ICRF/NHS Centre for Statistics in Medicine, Oxford †

Mr Peter Bower,
Independent Health Advisor, Newcastle-upon-Tyne †

Ms Christine Clark,
Honorary Research Pharmacist, Hope Hospital, Salford†

Professor David Cohen,
Professor of Health Economics, University of Glamorgan

Mr Barrie Dowdeswell,
Chief Executive, Royal Victoria Infirmary, Newcastle-upon-Tyne

Professor Martin Eccles,
Professor of Clinical Effectiveness, University of Newcastle-upon-Tyne †

Dr Mike Gill,
Director of Public Health and Health Policy, Brent & Harrow Health Authority †

Dr Jenny Hewison,
Senior Lecturer, Department of Psychology, University of Leeds †

Dr Michael Horlington,
Head of Corporate Licensing, Smith & Nephew Group Research Centre

Professor Sir Miles Irving
(Programme Director), Professor of Surgery, University of Manchester, Hope Hospital, Salford †

Professor Alison Kitson,
Director, Royal College of Nursing Institute †

Professor Martin Knapp,
Director, Personal Social Services Research Unit, London School of Economics & Political Science

Dr Donna Lamping,
Senior Lecturer, Department of Public Health, London School of Hygiene & Tropical Medicine †

Professor Theresa Marteau,
Director, Psychology & Genetics Research Group, UMDS, London

Professor Alan Maynard,
Professor of Economics, University of York †

Professor Sally McIntyre,
MRC Medical Sociology Unit, Glasgow

Professor Jon Nicholl,
Director, Medical Care Research Unit, University of Sheffield †

Professor Gillian Parker,
Nuffield Professor of Community Care, University of Leicester †

Dr Tim Peters,
Reader in Medical Statistics, Department of Social Medicine, University of Bristol †

Professor David Sackett,
Centre for Evidence Based Medicine, Oxford

Professor Martin Severs,
Professor in Elderly Health Care, Portsmouth University †

Dr David Spiegelhalter,
MRC Biostatistics Unit, Institute of Public Health, Cambridge

Dr Ala Szczepura,
Director, Centre for Health Services Studies, University of Warwick †

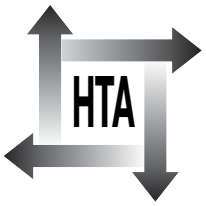
Professor Graham Watt,
Department of General Practice, Woodside Health Centre, Glasgow †

Professor David Williams,
Department of Clinical Engineering, University of Liverpool

Dr Mark Williams,
Public Health Physician, Bristol

Dr Jeremy Wyatt,
Senior Fellow, Health and Public Policy, School of Public Policy, University College, London †

* Previous Chair
† Current members



How to obtain copies of this and other HTA Programme reports.
An electronic version of this publication, in Adobe Acrobat format, is available for downloading free of charge for personal use from the HTA website (http://www.hta.ac.uk). A fully searchable CD-ROM is also available (see below).

Printed copies of HTA monographs cost £20 each (post and packing free in the UK) to both public **and** private sector purchasers from our Despatch Agents.

Non-UK purchasers will have to pay a small fee for post and packing. For European countries the cost is £2 per monograph and for the rest of the world £3 per monograph.

You can order HTA monographs from our Despatch Agents:

– fax (with **credit card** or **official purchase order**)
– post (with **credit card** or **official purchase order** or **cheque**)
– phone during office hours (**credit card** only).

Additionally the HTA website allows you **either** to pay securely by credit card **or** to print out your order and then post or fax it.

Contact details are as follows:

HTA Despatch
c/o Direct Mail Works Ltd
4 Oakwood Business Centre
Downley, HAVANT PO9 2NP, UK

Email: orders@hta.ac.uk
Tel: 02392 492 000
Fax: 02392 478 555
Fax from outside the UK: +44 2392 478 555

NHS libraries can subscribe free of charge. Public libraries can subscribe at a very reduced cost of £100 for each volume (normally comprising 30–40 titles). The commercial subscription rate is £300 per volume. Please see our website for details. Subscriptions can only be purchased for the current or forthcoming volume.

Payment methods

Paying by cheque
If you pay by cheque, the cheque must be in **pounds sterling**, made payable to *Direct Mail Works Ltd* and drawn on a bank with a UK address.

Paying by credit card
The following cards are accepted by phone, fax, post or via the website ordering pages: Delta, Eurocard, Mastercard, Solo, Switch and Visa. We advise against sending credit card details in a plain email.

Paying by official purchase order
You can post or fax these, but they must be from public bodies (i.e. NHS or universities) within the UK. We cannot at present accept purchase orders from commercial companies or from outside the UK.

How do I get a copy of *HTA on CD*?

Please use the form on the HTA website (www.hta.ac.uk/htacd.htm). Or contact Direct Mail Works (see contact details above) by email, post, fax or phone. *HTA on CD* is currently free of charge worldwide.

The website also provides information about the HTA Programme and lists the membership of the various committees.

Qualitative research methods in health technology assessment: a review of the literature

E Murphy
R Dingwall
D Greatbatch
S Parker
P Watson

School of Sociology and Social Policy, University of Nottingham, Nottingham, UK

Published December 1998

This report should be referenced as follows:

Murphy E, Dingwall R, Greatbatch D, Parker S, Watson P. Qualitative research methods in health technology assessment: a review of the literature. *Health Technol Assessment* 1998; **2**(16).

Health Technology Assessment is indexed in *Index Medicus*/MEDLINE and *Excerpta Medica*/EMBASE. Copies of the Executive Summaries are available from the NCCHTA web site (see overleaf).

NHS R&D HTA Programme

The overall aim of the NHS R&D Health Technology Assessment (HTA) programme is to ensure that high-quality research information on the costs, effectiveness and broader impact of health technologies is produced in the most efficient way for those who use, manage and work in the NHS. Research is undertaken in those areas where the evidence will lead to the greatest benefits to patients, either through improved patient outcomes or the most efficient use of NHS resources.

The Standing Group on Health Technology advises on national priorities for health technology assessment. Six advisory panels assist the Standing Group in identifying and prioritising projects. These priorities are then considered by the HTA Commissioning Board supported by the National Coordinating Centre for HTA (NCCHTA).

This report is one of a series covering acute care, diagnostics and imaging, methodology, pharmaceuticals, population screening, and primary and community care. It was identified as a priority by the Methodology Panel and funded as project number 93/44/04.

The views expressed in this publication are those of the authors and not necessarily those of the Standing Group, the Commissioning Board, the Panel members or the Department of Health. The editors wish to emphasise that funding and publication of this research by the NHS should not be taken as implicit support for the recommendations for policy contained herein. In particular, policy options in the area of screening will, in England, be considered by the National Screening Committee. This Committee, chaired by the Chief Medical Officer, will take into account the views expressed here, further available evidence and other relevant considerations.

Reviews in *Health Technology Assessment* are termed 'systematic' when the account of the search, appraisal and synthesis methods (to minimise biases and random errors) would, in theory, permit the replication of the review by others.

The editors have tried to ensure the accuracy of this report but cannot accept responsibility for any errors or omissions. They would like to thank the referees for their constructive comments on the draft document.

ISSN 1366-5278

Published by Core Research, Alton, on behalf of the NCCHTA.
Printed on acid-free paper in the UK by The Basingstoke Press, Basingstoke.

Copies of this report can be obtained from:

The National Coordinating Centre for Health Technology Assessment,
Mailpoint 728, Boldrewood,
University of Southampton,
Southampton, SO16 7PX, UK.
Fax: +44 (0) 1703 595 639 Email: hta@soton.ac.uk
http://www.soton.ac.uk/~hta

Contents

List of abbreviations

A&E	accident and emergency
CA	conversation analysis
CAQDAS	computer-assisted qualitative data analysis software
CCT	controlled clinical trial
GP	general practitioner
HIV	human immunodeficiency virus
HSR	health services research
HTA	health technology assessment
NUD*IST	Non-numerical Unstructured Data Indexing, Searching and Theorising
RCT	randomised controlled trial
SIDS	sudden infant death syndrome
VAMP	Value-Added Medical Products

Executive summary

Background

Qualitative research involves the collection, analysis and interpretation of data that are not easily reduced to numbers. These data relate to the social world and the concepts and behaviours of people within it. Qualitative research can be found in all social sciences and in the applied fields that derive from them, for example, research in health services, nursing and pharmacy. These research methods are not a recent invention but their application in health technology assessment (HTA) or health services research may be novel. In order for commissioners and researchers to utilise these methods and gain valuable knowledge from the results obtained, it is important that they understand the principles of qualitative methods and the way that they may be used to set benchmark standards.

Objective

The objective of this review was to examine the nature and status of qualitative methods in relation to their potential uses in HTA.

Methods

The search tools often used for systematic reviews were not appropriate for this review as it would be necessary to cover the equivalents of MEDLINE in a range of disciplines and applied fields, many of which do not have databases of comparable coverage. In addition, important methodological writing in the field of social science started long before indexing for computer databases, and much of the most significant work has been published in books rather than journals.

Having set the boundaries and organised the categories for this review, therefore, the authors read as widely and as comprehensively as was feasible in the time available. The authors have compared different researchers' approaches to the same issue and examined the work of authors who offer different perspectives.

Perspective

Idealists versus realists

Qualitative work is often identified with idealism while quantitative work is identified with realism. However, most qualitative researchers accept that there is an objective, material world, as do realists, but question our ability to know this directly. In the social sciences, what people perceive or believe is the basis of their actions rather than what an impartial observer might think was actually true.

Qualitative versus quantitative methods

The goal of all research in HTA should be to establish knowledge about which we can be reasonably confident, and to provide findings that are relevant to policy makers and practitioners. Therefore, decisions about whether qualitative or quantitative methods (or a combination of both) are most appropriate to a particular research problem should be made on the basis of which approach is likely to answer the question most effectively and efficiently.

Qualitative methods are useful in the exploratory stages of a research project, where they will often help to clarify or even set the research question, aid conceptualisation and generate hypotheses for later research. Qualitative methods may also be used to interpret, qualify or illuminate the findings of quantitative research and to test hypotheses.

Qualitative research is particularly useful to policy makers and planners by providing descriptive information and understanding of the context in which policies will be implemented.

Sampling and generalising

In sampling decisions in qualitative research, pragmatic considerations should be integrated with sampling in a systematic way just as in quantitative research; opportunistic sampling should

be avoided if possible. The emergent nature of qualitative research means that sampling decisions need to be made throughout the study; such decisions should again be systematic and principled.

Where the aim is to build or develop theory, subjects should be selected accordingly; such theoretical sampling makes use of existing theory to make predictions, and then seeks subjects who allow the researcher to test the robustness of such predictions under different conditions.

Methods of qualitative research

Participant observation

Participant observation can be used to study the impact of technologies upon the routine functioning of the setting in which they are to be implemented. Participant observation may also be used to review health technologies currently in practice, and has the potential for uncovering the process through which professional inputs are transformed into patient/client outcomes thereby identifying opportunities for modifying current practice to improve outcomes.

Interviews

Qualitative interview techniques are used, particularly in exploratory research, to study the range and complexity of ideas and definitions employed by individuals and groups involved in the implementation of health technologies. Both qualitative and quantitative interviewing share the same fundamental problem, however, in that they rely upon interviewees' reports and such reports are necessarily constrained by the context in which they are collected.

Written records

The analysis of written records has an important contribution to make to our understanding of the processes and consequences associated with new technologies. In addition, documents such as health diaries may provide important data on the implementation of health technologies.

Conversation analysis

The techniques of conversation analysis can provide detailed data on the impact of new technologies upon healthcare settings, the organisation of professional work and the interactions between health professionals and patients.

Research ethics

The same ethical principles apply to qualitative and quantitative research in HTA. The mechanical application of ethical codes developed in the context of biomedicine may be unduly constraining in qualitative research and may distract from those ethical risks which are specific to qualitative research. Covert research will rarely, if ever, be justified in HTA. Such research is likely to be a betrayal of trust and a gross invasion of privacy.

Assessment of qualitative research

The same assessment criteria of validity and relevance are appropriate for both qualitative and quantitative research in HTA.

Relevance

The relevance of HTA research is related to its potential generalisability to groups or settings beyond those studied. Given that most qualitative research is based on a single case or only a small number of subjects, the generalisability of qualitative research is achieved by the generation of theoretical statements, which may guide policy makers but remain to be tested through application in other contexts.

Data handling

HTA commissioners should look for evidence that applicants intend to use systematic methods for coding and handling their qualitative data and that methods proposed for analysing such data are appropriate to the research objective.

Computerised analysis packages for qualitative data offer an efficient way of handling qualitative data sets and may improve the rigour of the analysis by facilitating searches for falsifying evidence. However, such programs should be used only as a means of facilitating the analysis process rather than carrying out the analysis, which depends upon the theoretical sensitivity of the analyst.

Judgements about the validity of research depend upon being able to form a judgement of the research process. Researchers therefore need to provide a detailed record of their methods. Given the non-standardised nature of qualitative research, such records are likely to be more elaborate than in reports of quantitative research.

The trustworthiness of data analyses is enhanced where researchers can demonstrate that they have considered alternative plausible explanations for their data. The validity of research findings is enhanced where the researchers increase our understanding of all members in a setting and do not

present one-sided accounts. Confidence in the validity of findings is increased where there is evidence of researcher sensitivity to the ways in which the data have been shaped by the researchers' presence.

While the practices of respondent validation and triangulation may increase the comprehensiveness of a study, neither can be treated as tests of the validity findings.

Conclusion

There are strengths and limitations to qualitative approaches as there are to quantitative methods. However, where qualitative research is conducted properly and data analysed thoroughly, this approach can provide valuable information on the implementation and impact of health technologies on both health professionals and patients.

Preface

The objective of this review is to examine the nature and status of qualitative methods in relation to their potential uses in health technology assessment (HTA). The original guidance to the team from the commissioning panel was that **health technology** should be understood broadly to include human service innovations as much as new equipment or tools. In this sense, it anticipates the recent restructuring of the National R&D programme and some elements of the report would now find a more natural home in relation to the streams dealing with Service Delivery and Organisation or New and Emerging Technologies.

This is neither a systematic review in the sense of the 'York Model' developed by the NHS Centre for Reviews and Dissemination (CRD) nor is it an introductory guide to qualitative methods. The nation's bookshops are well-stocked with texts competing to offer the latter service, many of which are cited in this review. Our brief was to help research commissioners and users make appropriate purchasing and utilisation decisions rather than to produce training materials. The *Overview* (chapter 1) focuses on points which commissioners should bear in mind when considering whether and how to purchase or evaluate qualitative research. It is keyed to the rest of the report where a fuller discussion and justification of each point is located. Some users will wish to begin with the main text and to treat chapter 1 as an *aide-memoire*. Others will choose to start with chapter 1 and to follow its references back to the main text as they require.

It may be useful to set out the reasons for our rejection, with the endorsement of the commissioning panel and original referees, of the conventional systematic review model since these have implications for other review commissions and for the approach taken in this report. Some of the reasons are practical. From the beginning, the objective of reviewing the literature on qualitative methods was clearly a daunting task, which it was difficult to conceive of approaching in a way that could be exhaustive or comprehensive in the sense of the York Model. Qualitative research can be found in all the social sciences – economics, psychology, sociology, anthropology, politics, geography and law – and in the applied fields that derive from them such as health services research (HSR), nursing research, pharmacy practice research, social work research and so on. It has a history that goes back at least two thousand years. Despite its scale, this report barely scratches the surface and several human lifetimes would be required to read everything that might conceivably be relevant.

The available search tools offer limited assistance in this task. As other reviews in the HTA Methodology programme have found, even where a field is well covered by MEDLINE, it is difficult to extract articles that focus directly on methods. Here, it would be necessary to cover the equivalents of MEDLINE in a range of disciplines and applied fields, many of which do not have a database of comparable coverage. Moreover, in social science, important methodological writing has been going on for much longer than computer databases have been indexing and much of the most significant work has been published in books or edited collections rather than in journals. These limitations of the York Model were underlined by the small scale experiment that we conducted with it, in the area of evaluation research. The results, reported in chapter 7, confirmed our expectations.

Our other objections relate to the philosophy of method involved in the CRD approach. In fact, the writing of this review has become an example of the process that it seeks to describe. The CRD guidelines express what we shall describe as the classic approach of nineteenth century positivism and its twentieth century descendants. Their logic is hypothetico-deductive. The reviewer starts by formally defining a search strategy to produce certain results. In principle, all definitional decisions are taken in advance and all professional judgements are eliminated by objective scoring systems that allow all results to be fed into a single matrix, which can then be analysed by impersonal means. This approach works well under certain limiting conditions. The topic being studied must be in a state of what Kuhn (1962) called 'normal science' where there is a high degree of consensus on the definition of problems and methods, where there are accepted means of defining these operationally which lead to a standard use of keywords and where the results come in forms that can be treated as equivalent or converted into a common currency.

None of these conditions applied in the present review nor, indeed, in many fields for which reviews might be relevant. The definition, objectives, means and language of qualitative research are all contested. There is no standard template that can be applied and no common currency to establish equivalence. In a methodological review, there are not even empirical findings to compare: the central themes are epistemological debates about what we can know and how we can know it or claim to know it. As a result, our approach is essentially a qualitative one which exhibits the strategies of the kind of qualitative research that we ourselves do. The 'Nottingham Model', if we may call it that, uses a process of analytic induction. We begin from our own professional knowledge and judgement to sketch a field to survey and the likely dimensions for investigation. Having set our boundaries and organised our categories, we have then sought to read as widely and comprehensively as was feasible in the available time and to organise our reading around our own heuristic selection of key terms. We have compared different authors' approaches to the same issues (constant comparison), we have looked for authors who offer different perspectives (deviant case analysis) and we have tried to establish when there is no more to say on an issue (theoretical saturation).

Although the source of occasional frustration, we have also borne in mind that this is a review and not an occasion to argue a particular case. One of our proposed criteria for quality in qualitative research is fair dealing with viewpoints that are not the author's. In a contested field, it is not the job of a review like this to legislate for one version or another. This is not a *Which?* report, defining a 'best buy' in qualitative methods. As with all research, the Best Buy depends upon the starting question and the sorts of answer that are required.

We have, however, sought to indicate the dimensions along which the reader might evaluate the various arguments that we have summarised here. It is important, though, to stress that the criteria for evaluating writing about methods are not necessarily those for evaluating methods themselves. Empirical reports might be weighted on a scale depending upon their methodology – though it often seems perverse to rate a poorly-designed trial ahead of a well-conducted observational study purely to avoid making a judgement that the reviewer should be competent to make. The evaluation of arguments involves a different sort of assessment which will be influenced by what the reader considers persuasive and what the objective of the argument is intended to be. The methodological literature which bears directly upon the application of qualitative methods to HTA is, as yet, very limited. We have therefore conducted this review drawing upon the voluminous literature on qualitative methods in social science and health research. Where possible, we have drawn our examples from healthcare settings. A number of these examples recur throughout the review as they illustrate different points. Summaries of selected studies are presented in appendix 1. Chapters 2–5 of the report each conclude with a short summary of what we take to be the implications of this literature for the commissioning and practice of research in the field of HTA. Given the level of disagreement among qualitative researchers and the fact that such disagreement arises from differences in the *a priori* assumptions that such researchers make, it is inevitable that these summaries will reflect our own particular methodological position. However, this position has been developed through and subjected to the literature reviewed in the main body of the text. No doubt, however, since what is involved is a weighing up of arguments, different reviewers might have come to different conclusions.

In studying this text, then, readers are invited to turn its discussions of method back onto the work itself and to see it as the product of a qualitative research process. Like all qualitative studies, it is also a process that is conditioned by the circumstances of its production. Despite the scale of work represented here, it is the product of a team whose main knowledge and experience lie in sociology and, to a lesser extent, history and anthropology. Although we have attempted to cover qualitative methods elsewhere, the coverage in some areas could have been fuller with more time or resources. This is particularly true of psychology. In other areas, there seems to be too little secondary material and primary scholarship beyond the scope of the commission would have been needed. This applies particularly to the qualitative traditions in economics, though these have had little present-day impact on health economics. It might be argued, however, that a fuller treatment could have brought out the narrowness of the range of economic thought that is presently being deployed on the problems of health technology and health services. Our contention, though, is that this review is comprehensive in terms of **issues**, that there is no technique or debate of general importance that has been omitted, even if its precise impact on a particular discipline or field has not been fully explored. In making use of this report, commissioners should have a framework that can be applied to any of the disciplines or fields of study that they are concerned with.

The main report is divided into six chapters. Each of these sections is designed to stand independently of the other five, to allow selective reading. The sections are cross-referenced to one another to allow the reader to follow up ideas and issues which are treated more fully, or from a different angle, in other parts of the report. In chapter 2, qualitative methods are placed within their historical context. This shows how contemporary practice in qualitative research rests on a deep, but often invisible, foundation of methodological debate, which has elaborated and refined the justifications for studying social life in this way. It demonstrates that qualitative research is neither a novelty of the last decade nor necessarily the refuge of those who cannot think rigorously enough to do statistics. Chapter 3 is concerned with the relationship between qualitative and quantitative methods. In this section, we consider the basis upon which the choice between qualitative and quantitative approaches can be made, reviewing the range of positions which different authors have taken up on this issue. We then move on to consider the various dimensions, both philosophical and methodological, along which it is sometimes argued that qualitative and quantitative approaches differ in fundamental ways. In chapter 4, we turn to a consideration of the specific techniques which are used by qualitative researchers, and contextual issues, such as sampling, analysis and ethics. In chapter 5, we discuss criteria for assessing qualitative research. Chapters 6 and 7 contain case studies, which illustrate and discuss the application of qualitative methods to areas which are of particular relevance to HTA. In chapter 6, we illustrate many of the arguments reviewed in chapters 3–5, by discussing the use of qualitative and quantitative methods to study the use of medical computer information systems. Alternative qualitative approaches to this topic are discussed, drawing upon two exemplary studies. In chapter 7, we turn to the use of qualitative methods in programme evaluation. A review of the literature on qualitative and quantitative approaches to programme evaluation is followed by a detailed analysis and critical discussion of reports of healthcare evaluations published in seven selected journals in 1995.

Chapter 1

Qualitative research methods in health technology assessment: an overview

This chapter is intended to function as a statement of points for research commissioners to hold in mind when considering whether and how to purchase or evaluate qualitative work. It is keyed to the rest of the report, which provides a more extended treatment of the literature that lies behind any particular statement that is made here.

1.1 The scope of the review

Qualitative research can be found in all the social sciences – economics, psychology, sociology, anthropology, politics, geography and law – and in the applied fields that derive from them such as HSR, nursing research, pharmacy practice research, social work research and so on. This review has drawn most heavily on sociology and anthropology, where the debates have been most fully documented and has noted some gaps in coverage where secondary literature is either non-existent or difficult to locate. Nevertheless, the arguments involved are, in general, common to all these fields of study, with the main differences lying in the chronology.

1.1.1 Some fundamentals

The field of qualitative research is a highly contested one in which there are extensive disagreements about the nature, purpose, status and practice of its methods. It is essential that both commissioners and practitioners of qualitative research are clear about these differences if they are to identify and support the versions of such research that are most appropriate in HTA. As chapter 2 shows, the intellectual traditions on which qualitative research draws go back over two thousand years and represent a position in debates over fundamental questions of ontology, epistemology and methodology that are unlikely ever to be conclusively resolved, and which underpin current methodological debates.[1]

Detailed below are four crucial points of disagreement within the qualitative tradition to which commissioners need to be sensitive.

1.1.1.1 Is there a world out there?[2]

In contemporary terms this occurs as the question of social constructionism, though we see similar issues in the ideas of scepticism, of hyperbolic doubt, of immaterialism or of idealism. In essence it refers to the problem of what we can know about the real world apart from our perceptions of it. If we think it is all a question of perception, then, in principle, we can redefine the world in any way we choose. Researchers in this position will offer either an alternative perspective on an HTA problem, reflecting their own values or beliefs, or a summary of the multiple perspectives held by the various people involved in the activity being studied. There may be situations in which this is a valuable exercise in offering alternative ways of thinking about the organisation or delivery of a service or about its possible impact on employees or users. It may be a way to stimulate the imagination of managers or clinicians and to challenge accepted ways of thinking and working. As such, however, investment may need to be judged against the possible costs of, say, an artist in residence or a patient advocate.

Other qualitative researchers have developed procedural responses to this problem. While it may be conceded that, in some ultimate sense, all reality is the product of socially organised ways of perceiving it, it is recognised that the material world imposes a structure of its own and that the classifications of the natural sciences are fairly stable and resistant to change. While it may sometimes liberate the scientific imagination to stress the artificiality of, say, contemporary taxonomies of disease, the pragmatic achievements of clinical science justifiably make this a fairly rare event. It is, then, accepted that the natural sciences have a distinct method and logic of inquiry which can, most of the time, disregard the problematic relationship between perception and reality and function as if they had a direct access to an objective material world. When we come to social affairs, however, it is the organisation of perception that is important as the basis of social actions. The natural

[1] See chapter 2; [2] See sections 2.2 and 3.2.2.1.

scientific and social scientific questions about health technologies are just different and require different modes of investigation.

Where appropriate qualitative work is disciplined by the socially organised nature of perception. We cannot just arrange the world any way we please but are constrained by our membership in various kinds of group to perceive it in ways that other members will regard as sensible. It is, then, argued that there is an order and regularity to be found, particularly by concentrating on the typical, the everyday and the routine in a social setting and by aiming to connect the social scientist's analysis to the way that the people involved understand it themselves. The exotic or the scandalous incident is only useful inasmuch as it provides a contrast or control, which illuminates features that usually pass unnoticed because they are so stable and regular.

From this point of view, commissioners can obtain an understanding of how a social setting works and how particular outcomes are produced by the mundane activities of the people involved. Such an understanding can be a basis for prediction or, perhaps better, for a better-informed judgement about the likely effect of changes.

1.1.1.2 Induction and deduction[3]

Much contemporary qualitative work stresses its inductive character, while quantitative work tends to stress its deductive character. In fact, it is clear that good science involves both for different purposes at different times. Again, there is a commissioning judgement to be made about whether the objective is to get a feel for a particular case or to try to move towards a body of generalisations. Induction can be disciplined by the sort of canons that JS Mill lays out – the methods of agreement, difference, residues and concomitant variations. However, qualitative research can also be done in a deductive fashion, where prior theories or generalisations are tested on new cases. As these generalisations become better-founded, they can be used by managers or clinicians for practical goals.

1.1.1.3 Fact/value[4]

Although some qualitative research makes a commitment to a particular set of prior values and then uses its data to expound them, the same is true of some quantitative work. It is important to distinguish three different kinds of value issue. There is the study of means/end relations: if an NHS manager or clinician has adopted a particular policy on a particular normative basis, will the chosen means of operationalising it achieve the desired ends? There is the objective study of values: how and why does someone come to hold a particular ideal, belief or perspective on policy? Finally, there is the choice of values: which ideals, beliefs or policies **should** be adopted?

Most social scientists, qualitative or quantitative, would see a clear role in relation to the first and second of these but would argue that their place is to **inform** debate on the third. It is for policy makers, planners and managers to **decide** what should be done. Social scientists may have views on this as citizens but this should not necessarily give them priority over the views of other citizens. Having said this, the fact that they have carried out research into a topic, and spent time thinking about it, should mean that these views are better informed than those of most other citizens and it may be rational for a policy maker to take particular account of them. A commissioner should not, though, expect to pass on the responsibility of choice. However, some social scientists would claim that their studies can give them a privileged position in respect of ideals, beliefs and policies. It is important to separate two slightly different claims here. One is that some ends may be morally superior to others, which is associated particular with certain kinds of critical theory and standpoint research, where data are used to illustrate the normative claim. The second is that some ends may cut across fundamental principles of social organisation. It is, for example, argued by some that markets depend upon frameworks of values, which they cannot themselves generate. This may appear to offer a rational basis for moral action but in fact depends upon a prior choice, in the example that markets are to be preferred to other forms of social organisation, and is, in that sense, a variation on the first type of value-oriented research. If a policy maker has adopted a market approach, what are the costs and benefits of this and what are the means of operationalisation?

1.1.1.4 Other people's minds[5]

One of the endemic problems for all social science is the fact that our species is not known to be telepathic. If we accept that actions are the result of some kind of process by which either the material world or other social actions can be perceived, interpreted and responded to, then we have to gain access to that process in order to explain its observable outcomes. How can we do this? For a long time, the answer in qualitative research

[3] See sections 2.3.2 and 3.2.1.2; [4] See sections 2.3.1 and 3.1.3; [5] See sections 2.3.1, 2.4 and 3.2.2.1.

was by an imaginative act of sympathy, empathy or *verstehen* (interpretive understanding), where the researcher attempted to imagine him or herself into the position of the person under study. Although the German sociologist Max Weber attempted to discipline this approach by insisting that it must lead to consistent generalisations and to a search for independent verification, his caution has had a limited impact. More recently, however, some social scientists have rejected the *verstehende* concern for meaning in favour of an emphasis on practice. Rather than trying to get inside the heads of the people being studied, they have examined practices or sequences of behaviour and considered what kinds of generative process might be necessary to produce them. (This difference is illustrated in the case study in chapter 6) The important point for commissioners is to establish the extent to which the reconstruction of the motives, intentions or thought processes of the people being studied is subject to some kind of explicit discipline. This may be done through a clear display of the reasoning employed by the researcher, by the presentation of data in a form that allows the reasoning to be checked and verified by readers, by reference to other findings that can provide a framework of corroboration and by a systematic consideration of actual or possible disconfirmatory data.

1.1.2 The relationship between qualitative and quantitative research[6]

There are three main positions on the respective roles of qualitative and quantitative research.[7]

1.1.2.1 Qualitative research as a precursor to quantitative

Even among fairly hardline quantitative researchers, there is a broad measure of agreement that qualitative methods can be useful in a predesign stage of research. A qualitative study may help to define the dimensions that a quantitative study would aim to measure and to suggest effective ways of asking questions by describing the language used by or intelligible to the population being examined.

1.1.2.2 A pragmatic choice

Many social scientists see the choice of methods as related to the commissioners' original questions. What do they want to know? Some kinds of questions require a quantitative answer and the data are naturally made available in a numerical form. Other kinds of questions require a qualitative answer and the data are naturally qualitative. Neither should be treated as a 'gold standard' for the other. Qualitative methods are thought to be particularly useful where the commissioner has questions about processes rather than outcomes[8] or where quantitative research has thrown up puzzling results or identified deviant cases that seem to do much better or much worse than the population norm.[9] Qualitative methods are particularly suited to answering 'How does this come to happen?' questions rather than 'How many?', 'How much?' or 'How often?' questions.

1.1.2.3 Qualitative research as senior partner

Some social scientists argue that qualitative research should be treated as a gold standard for quantitative work because of its inherently more comprehensive approach and greater validity. The problem here is reconciling this claim with the suspicion that many of its advocates have about a realist epistemology,[10] which makes the very notion of a gold standard questionable because it implies a prior reality which can control the methods applied to it.

This uncertainty colours the two other positions that can be found, and which are discussed below.

1.1.2.4 Incommensurable paradigms[11]

This position asserts that qualitative and quantitative research offer such different ways of knowing about the world that they simply cannot be compared. Our grounds for choosing between them are essentially matters of personal preference and the virtue in sustaining both is to maximise diversity in ways of knowing which can stimulate different ideas or approaches. However, it may be difficult to use these ideas without some basis for deciding whether some are more valid than others.

1.1.2.5 Critical theory[12]

Finally, there are those who will argue that they should be allowed to use whatever method produces data and results that are most favourable to their prior normative position. This approach rejects most traditional ideas about science as a device to control practice in the interests of particular social groups.

Commissioners may wish to purchase research for any of these reasons. The important thing is to be clear about which is required. Is the objective to ensure that a potentially expensive survey will be

[6] See chapter 3; [7] See section 3.1.1; [8] See, for example, Silverman (1981), appendix 1; [9] See, for example, Gantley *et al,* (1993), appendix 1; [10] See section 3.2.2.1; [11] See section 3.1.2; [12] See section 3.1.3.

well-designed? Is the objective to find out more about an organisational problem that is not presented in a quantitative form or where the outcome data have identified a puzzle – why have two apparently similar clinics produced such different outcomes in patients, for example? Is the objective to stimulate debate and challenge complacency by forcing people to confront issues and arguments that would normally go unacknowledged?

1.1.3 The foundations of qualitative research[13]

Qualitative and quantitative research are often seen as mutually opposed kinds of work by advocates of each style. However, this review points to the diversity of positions within qualitative work and to the considerable degree of overlap with ideas in the quantitative tradition.

1.1.3.1 Idealism versus realism[14]

Qualitative work is often identified with idealism while quantitative work is identified with realism. In fact this is not a neat mapping. Most qualitative researchers accept that there is an objective, material world, as do realists, but question our ability to know this directly. What matters for the domain of social sciences is what people perceive or believe rather than what might be ultimately true. **Radical constructionists** argue that reality is only what we make it. **Social constructionists** argue that we are constrained to perceive and act upon reality in certain ways as a result of our membership in particular social or cultural groups who would sanction any departure from their shared definitions. **Subtle realists** accept that material reality can itself be a constraint on the possibility of definition. We can only perceive the world in ways which are in some sense consistent with the immanent organisation of that world.

1.1.3.2 Truth

Strong realists assume that it is possible to state objective truths about the material world. Science produces descriptions and explanations which are literally true. Almost all qualitative researchers would be sceptical about this claim. Science actually produces the best descriptions and explanations that it can in a particular historical context with the tools available to it. Radical constructionists argue, however, that truth is essentially a matter of personal belief: what **I** believe to be the case. Social constructionists tend to view truth as consensual: what **we** believe to be the case. Subtle realists, on the other hand, see truth as a regulative ideal. It may be a matter of convention but that convention should aim to be as well-founded as current techniques make possible. Disagreements should be referable back to an observer-independent world as a check on which is more plausible, rather than evaluated purely as a matter of consensus.

1.1.3.3 Induction versus deduction[15]

Deduction is often associated with quantitative methods and induction with qualitative. Again, this review has established the misleading character of such dichotomisation. All quantitative work depends at some remove on inductive thinking. Similarly, qualitative work can begin from and return to deductive thinking. While the sympathy for induction among qualitative researchers gives them a particular value in relation to hypothesis generation for subsequent quantitative work,[16] this should not obscure the potential value of deductive qualitative work, which seeks to compare and test the conclusions of qualitative studies conducted in other settings and generalised through the abstracting language of theory.[17] This is a growing tendency in qualitative work with the expansion of the body of research on which such comparative studies can be based.

1.1.3.4 Natural versus artificial[18]

It is commonly claimed that qualitative research is the study of natural settings while quantitative research studies artificial settings. This confuses a number of issues, not least by implicitly identifying quantitative work with experimental research. In fact, of course, quantitative research is often done in natural settings, as in ethological observations or even survey questionnaires taken into people's home. Conversely, it cannot be assumed that a natural setting is left unaffected by the presence of an observer or that an interview simply copies the natural responses of the interviewee to actual experiences. Qualitative research can also be experimental, introducing changes and then observing their effects.

Good research from whatever source will be aware of its founding assumptions. For a commissioner, the issue here would seem to be the need to recognise that the hesitancy of many qualitative researchers when dealing with claims to truth is a mark of their understanding of the complexity of the issues rather than a lack of certainty or confidence in

[13] See section 3.2; [14] See section 3.2.1.1; [15] See section 3.2.1.2; [16] See, for example, Gantley *et al,* (1993), appendix 1; [17] See, for example, Silverman *et al,* (1992), appendix 1; [18] See section 3.2.1.3.

their own work, provided that its practice bears public scrutiny in the ways discussed below.

1.1.4 The practice of qualitative research[19]

It has been suggested that six features are characteristic of qualitative research practice, though, as we have found, there is some degree of variation in relation to each of them.

1.1.4.1 Adopting the perspective of the people being studied[20]

While most qualitative researchers would endorse the view that human and social behaviour is **meaningful** behaviour, the implications of this position are highly contested among qualitative researchers.[21] For some, the meaningfulness of behaviour points to the principal strength of qualitative methods: the opportunity that such methods give to get close to those being studied so as to appreciate their perspectives and the meanings which inform their behaviours.[22] Quantitative techniques are criticised for riding roughshod over participants' meanings. For others, this characterisation of qualitative research as preoccupied with members' meanings is misleading.[23] They argue that a substantial body of qualitative research, particularly that which is informed by ethnomethodology, is not concerned so much with what people **think** as with what they **do**. Indeed, these critics raise significant problems about the feasibility of uncovering participants' meanings. First, these meanings may be inaccessible to the participants themselves. Second, any account which participants produce will be a situated account, which bears an indeterminate relationship to the participant's decision to act in a particular way.[24] Third, the researcher does not simply reproduce participants' meanings. Rather, it is inevitable that a process of selection and interpretation, informed by the researcher's theoretical framework, intervenes between the researcher's observation and the account which (s)he gives of that observation.

All of this suggests that the task of uncovering participants' meanings is not as straightforward as some qualitative researchers suggest. Commissioners will wish to know how researchers intend to grapple with the difficulties of interpreting others' meanings, and the disciplines that will be used to avoid anecdotalism and the imposition of the researcher's prior assumptions upon his or her observations.

1.1.4.2 Description of the setting of the study[25]

Much qualitative research is concerned with description. By focussing upon the mundane detail of what is going on in a setting, qualitative researchers are in a position to challenge routine, but unexamined, assumptions about familiar social settings.[26] Indeed some qualitative researchers see description, rather than explanation, as the central goal of qualitative research. Such authors frequently attempt to distinguish between social scientific description and everyday description, but it is not clear that this distinction is as fundamental as some suggest. An awareness of the value of description must be tempered by an awareness that description can never be treated as a direct reproduction of reality. Many different descriptions of a setting, all of which may be valid, are possible and descriptions are necessarily imbued with theoretical assumptions. Commissioners should expect researchers to make their theoretical assumptions explicit in their proposals and reports.

1.1.4.3 Emphasis on context and on holism[27]

One of the major strengths of qualitative research lies in its emphasis upon understanding the phenomenon of interest holistically. This leads to a style of research which, rather than isolating and manipulating variables, as in most quantitative research, seeks to study phenomena in context.[28] Whereas many quantitative researchers aim to isolate causal relationships from the context in which they occur, qualitative researchers eschew such **context-stripping** and place complexity at the centre of their research. Such attention to context is particularly relevant to HTA insofar as it offers a bridge between the discovery that particular patient, clinical and/or organisational management strategies are effective under experimental conditions, and the efficient implementation of such findings in clinical settings.

1.1.4.4 Emphasis on process[29]

The emphasis on process in much qualitative research marks it out from the input–output model, which is characteristic of much quantitative

[19] See section 3.2.2; [20] See section 3.2.2.1; [21] See chapter 6 for a comparison of studies that adopt very different positions in this respect; [22] See, for example, Morgan and Watkins (1988), appendix 1; [23] See, for example, Silverman *et al,* (1992), appendix 1; [24] See, for example, Stimson and Webb (1975), appendix 1; [25] See section 3.2.2.2; [26] See, for example, Strong (1979a), Bloor (1976), appendix 1; [27] See section 3.2.2.3; [28] See, for example, Strong (1979a), Buckholdt and Gubrium (1979), Dingwall and Murray (1983), appendix 1; [29] See section 3.2.2.4.

research. Qualitative researchers emphasise the processual and dynamic nature of social life. This is associated with a preference for longitudinal designs and for prolonged engagement in the setting under study. This capacity for providing valid process data, to complement reliable outcome data from more quantitative studies, can be seen as one of the major contributions that qualitative research can make to HTA. Quantitative studies may be capable of establishing a link between input and output. However, if one wishes to make changes, understanding **how** a given input is translated into a particular output will be important. It is here that qualitative approaches, such as participant observation, can make a significant contribution.[30]

1.1.4.5 Flexibility of design and reluctance to impose a priori *frameworks*[31]

Unlike most quantitative research, qualitative researchers are often reluctant to foreclose on the design of their study in advance of the data collection. They sometimes describe their design as **emergent** during the course of the study. This may cause problems for commissioners who, understandably, would prefer to know exactly what they are purchasing before commissioning a piece of research. Qualitative researchers' commitment to flexibility in research design is associated with a reluctance to impose prior assumptions on the people or settings being studied, with the attendant risks of distortion and ethnocentrism. Rather, they suggest, the phenomenon of interest must be **discovered** and the researcher must be flexible enough to respond to such discoveries, through a process of **progressive focussing**.

The extent to which such flexibility of research design is essential will vary from topic to topic. Increasingly, qualitative researchers appear ready, at least in fairly general terms, to define the research question, in advance of the data collection. As researchers become more concerned with the accumulation of knowledge, the findings of earlier work are increasingly being used as a springboard for further research. However, concepts and theories drawn from earlier work are likely to be held lightly, and to be subject to reformulation or rejection as the study progresses. Researchers can (and perhaps should) be able to describe the procedures which they will use to discipline the progressive focussing of their study. However, it would be unfortunate if commissioners were to require such a degree of pre-specification of research design that the characteristic strengths of qualitative methods for discovery and offering novel understandings of familiar phenomena were sacrificed.

1.1.4.6 Conclusion

The argument that qualitative and quantitative research should be seen as incommensurable paradigms, marked by dichotomies of philosophy and practice, is likely to obstruct the useful application of qualitative methods to HTA. It is more profitable to recognise the complementarity of quantitative and qualitative methods, acknowledging the particular strengths of the latter in terms of their capacity for studying socially meaningful behaviour, holistically, in context and with due attention to the dynamic, processual nature of social events and interactions. The emphasis, within qualitative research, on flexibility of design is particularly well suited to hypothesis generation and discovery. The reluctance to impose (as opposed to incorporate) prior theory has significant potential for undermining the taken-for-granted assumptions, which may distort the findings, or undermine the usefulness of research in the field of HTA.

1.1.5 Specific issues of method[32]

1.1.5.1 Selection and sampling[33]

Probability sampling is rarely used in qualitative research.[34] Sometimes, this reflects the view that the goal of qualitative research is to understand unique cases ideographically, rather than to permit public generalisation. However, many qualitative researchers use non-probability sampling for practical rather than principled reasons. Practical and resource constraints mean that random sampling methods are often both inefficient and expensive. The ratio of settings studied to the number in the population of interest is usually too low to permit statistical inference. Although qualitative research could, in principle, be carried out on larger samples, in most cases qualitative researchers are obliged by time and resource constraints to trade breadth for depth.

The difficulties of using probabilistic sampling methods do not mean that qualitative researchers have to settle for haphazard selection procedures.[35] Such researchers can be expected to present a clear rationale for the sampling procedures they use. Such a rationale may draw on one of two main arguments. On the one hand, qualitative

[30] See, for example, Silverman *et al,* (1992); Bloor (1976), appendix 1; [31] See section 3.2.2.5; [32] See chapter 4; [33] See section 4.1; [34] See section 4.1.1.1; [35] See section 4.1.1.2.

researchers may seek to achieve generalisability using non-probabilistic methods, demonstrating the typicality of the setting or settings to be studied. Here the researcher selects the setting on the basis that it is sufficiently similar to other settings to which generalisation is sought to permit case-to-case transfer. Alternatively the researcher may use some form of **theoretical sampling**.[36,37] Here the grounds for selection of a case are that it is expected to exhibit or test some identified theoretical issue. The researcher may use theory to make predictions and then use theoretical sampling to test the robustness of such predictions under different theoretical conditions. Such theoretical sampling may be carried out in advance of the data collection, making use of theory derived from earlier work. Alternatively, it may be ongoing throughout the course of a research project, as the researcher seeks to develop and refine theoretical propositions which emerge from the research. It is important that qualitative researchers are clear about whether their sampling strategies are directed towards empirical (albeit non-probabilistic) generalisation or theoretical generalisation.

In qualitative research, sampling decisions will often be made at two levels.[38] In the first place the researcher will have to make some initial decisions about the group(s) or setting(s) to be studied. Second, further decisions may have to be made about what or whom to study within such settings or groups. In particular, such within-case sampling decisions will relate to the different temporal patterns within a setting, the range of types of people within the setting and the different contexts within a setting. Sound research proposals and reports will show evidence that the researcher has (or has used) systematic strategies for both initial selection of cases for study and for sampling within the case, if this is appropriate.

1.1.5.2 Observation[39]

Observational research is central to the qualitative research tradition and a number of observational studies have been carried out in healthcare settings.[40] In such research the observer gathers data on daily life in the group or setting under study. There are parallels here with the anthropologist's attempt to understand non-Western societies. The observer adopts the stance of a **learner** who seeks to understand the culture of the organisation or group under study. In qualitative observational studies, the focus of the research is expected to narrow during the data collection process, an approach which is sometimes known as progressive focussing.[41] The qualitative observer seeks to avoid the early (some would say premature) definition of variables of interest, which is required in structured observational studies. Once again, this may cause some concern to commissioners, who are more familiar with research protocols which are fully specified in advance of data collection. Nevertheless, particularly in the hands of an experienced qualitative researcher, progressive focussing can make a significant contribution to both the validity and the relevance of the research.

A key strength of qualitative observation studies is their capacity for studying processes, within settings, rather than merely identifying a relationship between input and output variables, as in much quantitative observational work.

There are significant differences between such qualitative observational research and more structured observational work, based on the use of standardised observational schedules.[42] In the latter the researcher avoids, as far as possible, interaction with those being observed. Typically, the length of time the observer spends in the setting is very short. Members' behaviour is coded by the observer, using pre-specified categories. The researcher seeks to treat the observed behaviours independently of their wider social, temporal and historical context. In qualitative observational studies, the researcher is likely to spend extended periods of time in the setting, seeking to avoid premature definition of variables and to understand observed behaviour in terms of the participants' own categories and in relation to the wider contextual features of the setting.

There is considerable debate within the qualitative research community about the relative merits of observation and interviewing, with a number of authors decrying the recent neglect of observational methods and the domination of qualitative research by the depth interview.[43] Observation is a labour-intensive research method and makes relatively heavy demands on resources. However, a number of authors have pointed to the advantages of observation over interview data. In particular,

[36] See section 4.1.1.4; [37] See, for example, Silverman *et al,* (1992), Bloor (1994), Dingwall and Murray (1983), appendix 1; [38] See section 4.1.2; [39] See section 4.2; [40] See, for example, Strong (1979a), Silverman (1981, 1984), Dingwall and Murray (1983), appendix 1; [41] See, for example, Silverman (1981, 1984), appendix 1; [42] See section 4.2.1; [43] See section 4.2.2.

observation side-steps some of the difficulties raised when one attempts to treat interview responses as a guide to actual behaviour. While interviews may generate large amounts of data, which appear relevant to areas of behaviour of interest to both commissioner and researcher, they raise substantial problems of validity.

A variety of roles are open to researchers within the settings under study.[44] These range from one extreme where the researcher hides his or her true identity and fully participates in the setting, playing the role of a full member. At the other extreme, the researcher avoids all interaction with the people in the setting. Most qualitative researchers adopt an intermediate position between these two extremes. Within any given setting, the researcher may occupy different roles on different occasions and in relation to different members of the setting. The choice of observer role can be expected to have a significant impact upon the data obtained and particular roles have particular advantages and disadvantages. While there is no single perfect role, commissioners can expect researchers to have made an informed decision about the role(s) that they will adopt within the setting studied and to be able to defend that decision.

The involvement of the observer in the setting under study has given rise to concerns about the potential **reactivity** of observational studies. Such reactivity can be over-estimated. The prolonged engagement of the observer in the setting is likely to reduce such reactive effects. In any research setting, the need for members to carry on with the daily business of their lives is also likely to reduce reactivity. The researchers should be expected to take such reactive effects into account in the analysis of their data. Indeed, the impact of the researcher on the setting under study may, in itself, represent useful data.

There is considerable debate about the relative merits of **covert** (secret) and **overt** observation.[45] Some defend covert research on the grounds that it reduces reactivity and gives access to settings which might otherwise be closed to research. However, it is likely that the inaccessibility of sensitive settings is exaggerated and constraints on what a covert researcher may legitimately do or observe within a setting may be of greater concern. It seems unlikely that commissioners would wish to fund covert research in healthcare settings. However, it is important to note that the line between covert and overt research is not a clear as is sometimes supposed. Even where researchers intend to be entirely open about their research activities, members of a given setting may have difficulty in fully grasping the purpose of the research.

1.1.5.3 Interviewing[46]

Researchers who use qualitative interviews justify their choice in relation to the perceived shortcomings of more structured, standardised interviews.[47] The latter are criticised for failing to recognise that all interviews are situated encounters between the interviewer and interviewee. The implication of this is that such interviews cannot be treated as uniform presentations of identical stimuli to all respondents. Such interviews mistake the standardisation of questions for the standardisation of meaning to respondents. Standardised interviews are also criticised for divorcing interview responses from the context in which they were produced and, on political grounds, for imposing a hierarchical and exploitative relationship upon respondents.

These deficiencies of standardised interviews are juxtaposed with the perceived virtues of qualitative interviews. Advantages proposed for qualitative interviewing techniques include the capacity for accessing respondents' definitions and interpretations of the phenomena under study[48,49] and for penetrating so-called **public accounts**.[50] The flexibility of qualitative approaches are seen as a particular strength, particularly in hypothesis generating studies.

Not all qualitative researchers share this enthusiasm for qualitative interviews.[51] Some argue that advocates of qualitative interviews face a set of problems which are essentially the same as those faced in more structured interviews. Both structured and less structured interviews must be seen as social situations. All interviews will be opportunities for **impression management** by both interviewer and interviewee. As a result, interviews, whether qualitative or quantitative cannot be treated as more or less accurate reports on some external reality. Rather they are occasions upon which respondents are called to account for their actions, feelings, opinions and so on. This means that what is said in an interview must be treated, not as a literal description to be evaluated in terms of the likelihood that the respondent is telling the

[44] See section 4.2.3; [45] See section 4.2.5; [46] See section 4.3; [47] See section 4.3.2; [48] See, for example, Morgan and Watkins (1988), Bloor (1994), Charmaz (1983b), appendix 1; [49] See section 4.3.3.1; [50] See section 4.3.3.2; [51] See section 4.3.6.

truth, but as accounts which are embedded in the circumstances of their production.[52]

1.1.5.4 Documents[53]

Textual and documentary sources of data have been relatively neglected in both qualitative and quantitative research. This is significant omission, particularly in the context of healthcare settings where written documentation is a routine and pervasive activity. Available documentary sources will include both informal documents (such as diaries, autobiographical accounts and letters), formal or official documents (such as administrative records, codes of practice, rotas etc.) and visual documents (such as photographs, advertisements etc.).[54]

Three rather different approaches to documentary analysis have been identified.[55] In **quantitative content analysis**, documents are treated as objective means for making inferences about the realities they record.[56] The concern here is with the manifest meaning of the documents, rather than uncovering deeper levels of meaning and practitioners are unconcerned with the role which such documents may play in maintaining existing power relationships. In **interpretative documentary analysis**, such documents are treated as significant social mechanisms which should be analysed in their own right, rather than as reflecting independent realities.[57] In the **critical tradition**, the focus is upon the role which official documents play in regulating the actions of particular groups within society.[58]

The status given to documentary evidence varies in qualitative research.[59] At times they are treated as potentially accurate records of the phenomena they represent. At others, they are analysed as accounts,[60] which are the product of the context in which they were generated.

Conversation analysis[61]

Conversation analysis (CA) involves the detailed analysis of audio and audio-visual recordings of naturally occurring social interaction to identify the interactional practices used by speakers to produce their own conduct and to interpret and deal with the conduct of others.[62] CA relies upon the procedures of analytic induction.[63] Increasingly it is being used used to analyse interactions in institutional settings such as medical consultations, explicating the taken-for-granted practices through which health professionals and patients manage their interactions, and to examine the ways in which people interact with and use technology.[64] A detailed discussion of the application of the techniques of CA within healthcare settings can be found in chapter 6.

1.1.5.6 Analysis[65]

The process of data analysis in qualitative research may, superficially at least, appear very different to quantitative analysis. In particular, qualitative analysis tends to be iterative rather than sequential. Whereas quantitative analysts seek to reduce data, generating a series of statistics, qualitative analysts are concerned to represent their data to their readers using detailed knowledge of cases. In qualitative analysis **theoretical sensitivity** is an important part of the analysis process. In qualitative analysis, it is more usual for the analysis to use coding categories that have emerged from the data, rather than those which reflect prior theoretical assumptions.

The end of point of qualitative analysis varies.[66] In some studies, the aim is to produce description. In others the aim is to produce, refine or test theory. By contrast to quantitative research, in qualitative research the processes of theory development and empirical data collection are likely to be interwoven. Two closely related approaches to theory generation/testing are frequently referred to by qualitative researchers: **analytic induction**[67] and **grounded theory**.[68] There are, however, relatively few studies which actually follow the highly demanding procedures of these methods of analysis. Commissioners will want to be reassured that, where researchers claim to be using such methods, they are clear about what is, in fact, involved and able to justify the choice of a particular method.

Qualitative research studies typically produce very large amounts of data and it is important that researchers should be able to demonstrate that they have strategies in place for managing such potentially unwieldy data sets.[69] One increasingly

[52] See, for example, Stimson and Webb (1975), Voysey (1975), Baruch (1981), appendix 1; [53] See section 4.4; [54] See section 4.4.1; [55] See section 4.4.2; [56] See section 4.4.2.1; [57] See section 4.4.2.2; [58] See section 4.4.2.3; [59] See section 4.4.3; [60] See section 4.3.6; [61] See section 4.5; [62] Given the variety of types of talk which have been studied and the increasing interest in studies of gesture and body movement, the title conversation analysis is increasingly inappropriate but is so well-established as to be effectively immovable. [63] See section 4.6.1.3; [64] See, for example, Heritage and Sefi (1992), Silverman *et al,* (1992), appendix 1; [65] See section 4.5; [66] See sections 4.5.1.1 and 4.5.1.2; [67] See section 4.5.1.3; [68] See section 4.5.1.4; [69] See section 4.5.2.

popular strategy involves the use of computer packages for the analysis of qualitative data. Such packages can indeed improve the efficiency of data management and may facilitate team working and multi-site projects. However, the use of such programs is not without risks. For example, their use may distance the analyst from the raw data, blunting the theoretical sensitivity, which is central to the process of data analysis in qualitative research.

The researcher wishing to use a computerised package to assist in the analysis of qualitative data is faced with a bewildering variety. There are fundamental design differences between the various packages and it is important that the choice of package should reflect the logic which underlies the proposed analysis. In a research proposal, researchers should be able to make the reasons for their choice of program explicit. These packages are not always easy to use and commissioners may wish to reassure themselves that the research team already has the necessary expertise or has made provision for acquiring that expertise in both financial and time budgets.

1.1.6 Ethical issues in qualitative research[70]

There are considerable problems with any attempt to translate ethical codes of practice directly from biomedical research to social scientific and, in particular, qualitative research. This is not because qualitative research does not pose a threat to the well-being of those who take part in it.[71] Rather it is because the risks that qualitative research does pose are different in kind to those of biomedical research. The extrapolation of codes of practice from the biomedical research to qualitative research may not only be unduly constraining to the latter, but may also distract attention from the kinds of harm which may arise specifically in relation to qualitative research.

In some cases qualitative studies could jeopardise participants' psychological or emotional well being.[72] Whereas, in biomedical research the greatest risk to the participant occurs during the research study, in qualitative research the most serious risks arise during the dissemination phase.

Qualitative research also raises particular issues in relation to the principle of informed consent.[73] Insofar as such research does not require subjects to abdicate control over the research process, it may be less of an issue than in biomedical research. On the other hand, the relative unpredictability of qualitative research design may mean that fully informed consent is impractical in advance of the research study.

An attempt at mechanical extrapolation of ethical safeguards derived from biomedical research is likely to be unhelpful. The same ethical principles (autonomy, protection of the participant from harm, justice) apply in all research involving human subjects. However, in qualitative research these cannot be guaranteed simply by requiring all participants to sign a consent form or weighing anticipated risks against anticipated benefits. Rather what is required is for the researcher to take responsibility for reflecting upon the possible ethical implications of a particular study and considering whether, in the light of these, the research should proceed. Where risks have been identified, the researcher should consider the means by which such risks may be reduced or eliminated. Finally the researcher must consider how the principle of informed consent may be observed, not only at the start of the study, but throughout the study as the research design emerges and the focus of the research narrows.

1.1.7 Assessing qualitative research[74]

1.1.7.1 The (im)possibility of criteria?[75]

The question of how (and indeed whether) qualitative research should be evaluated is highly contested within qualitative research.[76] Some argue that any attempt to apply criteria is doomed to failure because the very idea of criteria is incompatible with the anti-realist assumptions held to be foundational to qualitative research. We suggest that this position is unnecessarily constraining and severely limits the usefulness of such research in relation to HTA. Other researchers, again starting from an anti-realist position, argue that, since qualitative and quantitative research are grounded in fundamentally different paradigms, conventional criteria, such as validity and reliability are inappropriate to qualitative research.[77] If we adopt this relativist position, it is doubtful whether health technology can be assessed in any meaningful sense at all, whether by qualitative or by quantitative methods. If the findings of research cannot be taken to represent even an approximation to truth, it is not clear why commissioners should invest public money in funding such research.

[70] See section 4.6; [71] See section 4.6; [72] See section 4.6.2; [73] See section 4.6.3; [74] See chapter 5; [75] See sections 5.1 and 5.2; [76] See section 5.1; [77] See section 5.2.1.

The practice of qualitative research is not as inextricably linked to anti-realism as some would have us believe.[78] If we adopt a subtle realist position we are able to hold on to truth as a regulative ideal, while, at the same time, accepting that it will always be impossible to be absolutely certain that truth has been attained in any particular instance. This allows us to assess both qualitative and quantitative research in terms of two fundamental criteria – those of validity and relevance.[79] If research studies are to be used to inform the development and application of health technology then we must have some confidence that their findings are true. Likewise their relevance to the concerns of commissioners must be clear.

While these two criteria of validity and relevance apply to both qualitative and quantitative research, the substantial differences between the problems addressed and the methods used in the two traditions, mean that the methods by which error can be limited and relevance can be established will vary.

1.1.7.2 Validity[80]

It is neither possible nor helpful to offer a rigid checklist of rules, which qualitative research must observe if it is to be judged valid. The validity of a piece of research does not lie in the slavish application of rules or the use of algorithmic criteria. The risk of such checklists is that they become an end in themselves rather than enhancing validity. Even if the production of such a checklist were deemed wise, it is not clear what items should be included. Respondent validation[81] and/or triangulation[82] are often put forward as candidates for such a checklist but both raise serious problems as tests of validity.

Respondent validation or **member checking** involves presenting the researcher's analysis to research subjects for feedback on the validity of the conclusions. Such exercises are certainly capable of generating useful new data, but as tests of validity they raise significant problems. First, the accounts produced by researchers and research subjects are formulated in the light of very different purposes and can be expected to differ from one another in ways which have no bearing on their validity. Second, it may simply not be possible to persuade research participants to read research reports with the degree of attention that would be necessary for such a validation exercise. Third, it is impossible to disentangle participants' responses to the analyses from their situated behaviour in the validation exercise. Fourth, one cannot assume that participants will act as unbiased assessors in commenting on draft reports. They are likely to have their own agendas, which will be reflected in their responses. Finally, there is no guarantee that informants' responses will be consistent across time, even within a single interview. Viewed in the light of these criticisms, member checking is more appropriately seen as a further opportunity to search for negative evidence, rather than a validation exercise in the true sense of the term.

Triangulation involves the combination of methods in the study of the same phenomenon. The use of multiple methods of data collection is characteristic of much qualitative research and is, indeed, a major strength, insofar as it enhances the completeness of the data. However, some would go further than this and argue that triangulation can be used as a test of validity. This is the original sense of the term. The emphasis is upon counter-balancing the distorting effects of any single approach and the aim is to establish the convergent validity of findings drawn from complementary approaches. There are significant problems in seeking to treat triangulation as a test of validity. First, using triangulation as a test of validity rides roughshod over the commitment to analysing data in context, which is at the heart of qualitative research. Second, it encourages researchers to focus on a single 'master reality' rather than uncovering the situated work, which different versions perform in different context. Third, just because data from two different sources agrees does not mean that the researcher's inferences are valid. Both sources may be subject to systematic or random error. Fourth, there is a fundamental problem with a test of validity that can only corroborate findings and never refute them. This is the case with triangulation. If evidence from two sources is in conflict, we cannot know whether the differences arise from the shortcomings of one of the methods.

There are, then, very significant problems with adopting triangulation as a test of validity. Using a range of methods to investigate a problem may be useful in itself, particularly where the methods are chosen on theoretical grounds. Such mixed method approaches may add to the comprehensiveness of a study and stimulate reflexive analysis. In such cases the decision to use multiple methods will be guided by the availability of resources.

[78] See section 5.2.2; [79] See section 5.2.3; [80] See section 5.3; [81] See section 5.3.1; [82] See section 5.3.2.

There are, however, a number of practices, which serve to enhance the validity of qualitative research and to enable readers to draw conclusions about the trustworthiness of findings. It is particularly important that qualitative researchers should give a clear exposition of the way in which their data have been collected.[83] The researcher must provide a detailed description of the process by which the data on which the analysis was based were collected. Such careful description allows the reader to exercise 'joint responsibility' with the researcher in judging the evidence on which claims are based.

It is equally important that the analyst should describe the process by which their findings were derived from their data.[84] This will involve the clarification of the concepts and categories used in the research and the demonstration that the conclusions are justified in relation to the data collected. It is important that researchers display enough data to allow the reader to assess whether the analyst's interpretations are supported by the data. Readers will be more convinced of the validity of the analysis where alternative plausible explanations have been considered.

Qualitative researchers emphasise the embeddedness of research data in the circumstances of their production.[85] The analysis of research data should therefore involve careful reflection upon the ways in which the data have been shaped by the research process itself.[86] Such **reflexivity** will also take account of the researcher's own prior personal, theoretical and even political biases, recognising the role of values and *a priori* assumptions in shaping any research account.

The credibility of research reports is strengthened where researchers demonstrate that they have engaged in a conscientious search for data that are inconsistent with the emerging analysis.[87] The careful search for such deviant or negative cases allows researchers to refine their analyses until they can incorporate all available data. Claims to validity are strengthened where researchers display negative cases in their reports and then show how these can be integrated into the analyses. As in all science, it is the careful search for falsifying evidence which lends weight to the truth claims of qualitative research.[88]

This search for negative evidence can extend beyond the confines of an individual study. One criterion that may be used to judge qualitative research is the extent to which researchers have taken previous knowledge in their field into account. In some cases, researchers may seek to test the limits of the conclusions of a previous study in a new setting or among a new group.[89]

The commitment to **multiple perspectives**, which is characteristic of most qualitative research, has serious implications for the truth claims of any piece of research.[90] In particular, the researcher should be wary of presenting the perspective of one group as if this defined some objective truth about the phenomenon under study. Such fair dealing should extend not only to the relatively powerless within the setting, but also to the relatively powerful. Traditionally qualitative research has been concerned with the under-dog, but this risks failing to develop an even-handed analysis drawn from people at all levels of an organisation. The credibility and usefulness of a research report will be enhanced where the analyst conveys as much understanding of the powerful as of the powerless within the organisation.

Second criterion: relevance[91]

The issue of relevance is clearly central to any judgement of qualitative research in relation to HTA. However, one must be wary of defining relevance in terms that are unduly narrow. Useful research will be directed not only at the issues that are identified as being of immediate concern by practitioners, but also to those which the researcher can demonstrate **ought** to be of concern. It is often unhelpful to restrict research to those problems that are identified by practitioners or managers. We cannot always be sure that the problem presented by practitioners is in fact the one which needs to be addressed.

Relevant research will contribute to the cumulation of knowledge in a field rather than existing in magnificent isolation. It will add something to our existing knowledge. This will include confirmatory studies, studies that plug the gaps in what we already know, and studies that call into question what we thought to be self-evident.

One of the key issues that must be addressed in establishing the relevance of qualitative research concerns the extent to which findings can be generalised beyond the circumstances in which they were produced.[92] The combination of small sample

[83] See section 5.3.3; [84] See section 5.3.4; [85] See section 5.3.5; [86] See, for example, Baruch (1981) and Voysey (1975), appendix 1; [87] See section 5.3.6; [88] See, for example, Silverman (1981) and Strong (1979a), appendix 1; [89] See, for example, Dingwall and Murray (1983), appendix 1; [90] See section 5.3.7; [91] See section 5.4; [92] See section 4.1.

sizes and difficulties in using probability sampling methods in qualitative studies mean that statistical representativeness is rarely achieved in qualitative research. This does not mean that all qualitative researchers give up on the possibility of generalisation from their research. As discussed above, they may aspire to establish the empirical generalisability of their findings using non-probability methods to demonstrate the likelihood that the setting or settings studied are representative of the population to which generalisation is sought. Alternatively, they may seek to establish the theoretical generalisability of their research. Here the researcher may make use of theoretical sampling, analytic induction and the pursuit of negative cases, all of which are discussed in section 4.1.

1.2 Conclusion

Qualitative research is a highly contested field. Not only is it subject to criticisms from those who operate within alternative research traditions, there are also major cleavages among qualitative researchers on matters of ontology and epistemology as well as the methods to be used and the criteria which are appropriate for the evaluation of its products. The debates in relation to each of these aspects are fully discussed in the rest of the report. The hallmarks of high quality qualitative research are to be found in the same commitment to rigour, clarity and systematicity, which are the hallmarks of all good science.

Chapter 2

Qualitative methods: history and diversity

2.1 Where do we start?

Qualitative research methods are often presented in the context of research on health technology and health services as if they were something new, relatively untested and, hence, intrinsically controversial. In fact, of course, they are as old as human civilisation itself. When our ancestors first emerged on the African savannah, they would have studied the world around them using their native senses in precisely the same way as we do today. If we want to find out what another person, or a group of people, are doing or thinking, we use our eyes and our ears to watch and to listen and our voices to ask questions. If we want to find information at a distance, then we can send someone to collect observations or to ask on our behalf, or we can talk to those who have come from that other place or look at the ways in which they comport themselves. Since we have become a numerate and a literate species, we can do some of these things more efficiently. Some of the earliest surviving records are of national or household accounts and of tax levies (Shaw and Miles, 1979:28–9). Even civilisations that did not develop writing, like that of the Incas, created a means of recording financial data, in that case by an elaborate and now lost system of knotted cords, the *quipus*. Most agricultural societies devised a means of recording their cattle holdings, usually based on the 'natural system' of the human body with a base ten, or twenty, from the digits of the fingers and, sometimes, toes. However, when it came to the quality of those same cattle, they tended to develop elaborate vocabularies, permitting finely detailed descriptions of the age, sex, breed and markings such that any other user of the language could easily recognise individual animals (e.g. see Evans-Pritchard, 1940:41–8).

If qualitative research is the original and still endemic means by which most people find out about the world around them, it is, of course, difficult to delineate a precise starting point for an historical account. In one sense, we might begin with the written record of the Classical World where philosophers like Aristotle sought to make general statements about human nature on the basis of their own observations and the reports that they had collected from others. The examples could be multiplied: Herodotus, in his *Histories* and Thucydides in *The Peloponnesian Wars* developed methodological styles that are still familiar. They represent the break between a bard like Homer recounting a legend and a scholar seeking to write history (Collingwood, 1946:18). Both historians distinguished between the evidence of their own observations and interviews, their use of previous writers and their use of material evidence, like the inspection of ruins or inscriptions. They showed a degree of self-consciousness about the quality of their respective sources: Herodotus, for example, distinguished between his own **observations** of Egyptian life in the Nile Delta and the **reports** that he received about life up-river (Hornblower, 1987; Hunter, 1982; Lateiner, 1989). Whether this is merely a rhetorical trick or a substantive difference in Herodotus's work, which is a lively argument among classical scholars (Fehling, 1989), the idea that such a distinction could and should be made is of fundamental importance to writing about the social world. However, the distinction is simply made rather than analysed, and stands as a model that others may adopt if they happen to notice it rather than being precisely defined and justified.

This tradition continued to inspire scholars for a good 2000 years without fundamental change in its attention to questions of methodology. It would be largely irrelevant to the present exercise to trace this in detail. What we need to focus on is the point at which methodological discussion appears and writers feel obliged to produce justifications for the data that they have used to support their arguments. As we search for such a point, however, it seems to slither yet further away. In fact, there is no one history of self-conscious methodological writing. Each of the major social sciences has its own experience of a slightly different character over a different time period, which forms a subset of the emergence of the modern division of labour in the social sciences. Indeed, in certain respects, the dissection of the broad eighteenth and early nineteenth century vision of a human science, which would reach from the macro questions of political economy through the micro questions of individual reasoning and moral character to the biological determinants of both, may be as much the story of successive methodological disagreements as of theoretical ones.

For narrative convenience, however, we can divide the story roughly into four parts. The first part of this chapter (section 2.2) looks at what we might call the protohistory of social science, the emergence of a concern for how the world, broadly conceived, might be known and explained. This concern is identified with a series of scholars who might be more typically thought of as philosophers. However, as Ayer (in Dunn *et al,* 1992:206) has noted: "In reviewing the work of any of the famous seventeenth or early eighteenth-century philosophers, one must always bear in mind that they did not draw the distinction, which has more recently arisen, between philosophy and the natural or social sciences ... they regarded every form of scientific enquiry as philosophical".

This early story, from roughly 1600 until 1800, will be told with a fair amount of attention to the philosophical arguments, partly to substantiate the breadth of interests among these writers and partly to provide a foundation for the discussion of contemporary issues in the philosophy of social science in later chapters. It is in this period, for example, that we first find debates about whether we know the world around us as the creation of our perceptions or as an objective, observer-independent reality, which anticipate the current arguments around social constructionism.[1] These scholars also began to explore the respective merits of inductive and deductive reasoning[2] and to consider the nature of causal explanation in social matters.[3]

Section 2.3 explores two directions that branched off from this over the next hundred years or so. It begins by looking at the developments on mainland Europe and the version of social science that developed there. This part concludes by examining the *Methodenstreit* (literally the quarrel about methods) in Germany, which had a great influence on sociology, economics, psychology and history. This is then contrasted with the foundations of social science in the English-speaking world and the greater influence of positivist ideas.[4]

Section 2.4 overlaps with this but the focus is mainly on the professionalisation of social science and the development of methodologically self-conscious academic disciplines. There is a particular focus here on anthropology and the invention of participant observation as the pre-eminent means of coming to know about the life of non-Western societies. This method was then brought home to developed countries. A brief mention is made of developments in psychology, with the conflict between individual behaviourism and social behaviourism, and in economics, with the rise of the neo-classical paradigm and the retreat of institutional and Austrian approaches. However, it has not been possible to explore these in detail within the constraints of the present review. The section ends with a brief survey of the recent history of qualitative research in medical settings in Britain. This part deals less directly with the philosophical foundations of methodology, partly because these are covered elsewhere in the review for the contemporary period, and partly because there is a more explicit intermediary body of methodological writing.[5]

2.2 The invention of methodology

In medieval Europe, the principal authority for knowledge was the word of God as revealed through the teaching of the Roman Church or, just possibly, in the statements of eternal verity by those Classical scholars whose writings were approved by the church. Galenic medicine, for example, was an essentially static system, where Galen's account, or what purported to be the account of a man named Galen, described a fixed world, which had been unchanged since the moment of Creation. The power of reason was subordinate to the power of revelation. For a considerable period the Church opposed the notion that these accounts of an unchanging and unchanged Creation could be challenged by rational human inquiry:

> The Church's hostility rested ... upon what one might call the politics of epistemology. It could see all too clearly that any claim it might hitherto make to a privileged access to the ultimate, divine final cause would be fatally weakened by an alternative institution, science, being able to reveal that divine purpose in nature by describing nature in what we would now call straightforwardly naturalistic ways. (Hawthorn, 1976:9.)

As this quotation suggests, the Church's position did not remain unchallenged and there is an important body of medieval scholarship that looked for ways around its orthodoxy. There is also a great deal of debate among historians of science and medicine about the continuing influence of theological, mystical and astrological elements in Renaissance science. However, the break-up of the Church's authority in the sixteenth and

[1] See section 3.2.1.1; [2] See section 3.2.1.2; [3] See section 3.6.1; [4] See section 5.2; [5] See sections 4.2–4.6.

seventeenth centuries is so widely accepted as the beginning of what we call the modern world that it seems a sensible place to begin a more detailed discussion and to focus on the themes that have shaped our present understanding rather than those that look back to the past.

2.2.1 Scepticism

The Renaissance and the Reformation between them opened a space for reason, supplemented by the tools of experience and experiment, in understanding the world of nature, including the world of human nature. Michel de Montaigne (1533–92) and Francis Bacon (1561–1626) can be taken to represent the generation that rediscovered scepticism. As Bacon put it: "No-one has yet been found so firm of mind and purpose as resolutely to compel himself to sweep away all theories and common notions, and to apply the understanding, thus made fair and even, to a fresh examination of particulars" (Bacon, 1858:93).

More pithily perhaps, Montaigne adopted the phrase *Que sais-je?* (What do I know as a fact?) as his personal motto (Burke, 1981). In rejecting the dogmas of the medieval church, these writers began to ask how it was that we could come to know anything? If we could not believe the assertions of the past, by what means could we come to better conclusions about the world? In the *Novum Organum* (1608–20), Bacon laid out his programme. We would begin by experiencing nature and experimenting with her: "There remains simple experience; which, if taken as it comes, is called accident; if sought for, experiment...the true method [commences] with experience duly ordered and digested, not bungling or erratic, and from it educing axioms, and from established axioms, again new experiments..." (Bacon, 1858:81).

As the quotation shows, the word 'experiment' does not have its narrow contemporary sense, in the way that we might refer to an investigation conducted under controlled conditions. It is, rather, a deliberate seeking out of an occasion to test the validity of some proposition or axiom. In this respect, it also embraces the idea of the 'natural experiment' in contemporary qualitative work, the notion that generalisations may be tested by looking systematically for fieldwork sites that vary in specified ways that are believed to be relevant to the sustainability of the conclusions that have been drawn.[6]

From our experiences and experiments Bacon proposed that we could draw up **tables of discovery** – lists of the particularities of our observations – which we would then organise by building up from small generalisations to large ones:

> The understanding must not however be allowed to jump and fly from particulars to remote axioms and of almost the highest generality (such as the first principles, as they are called, of arts and things), and taking stand upon them as truths that cannot be shaken, proceed to prove and frame the middle axioms by reference to them; which has been the practice hitherto ... we hope well of the sciences, when in a just scale of ascent, and by successive steps not interrupted or broken, we rise from particulars to lesser axioms; and then to middle axioms, one above the other; and last of all to the most general. For the lowest axioms differ but slightly from bare experience, while the highest and most general (which we now have) are notional and abstract and without solidity. But the middle are the true and living axioms, on which depend the fortunes and affairs of men ... (Bacon, 1858:97.)

He rejected the idea that induction would proceed by simple enumeration, which he described as childish. Axioms are not developed by counting the numbers of instances or the numbers of particularities but by specifying logical relationships: induction "must analyse nature by proper rejections and exclusions". To the extent that induction moves beyond a particular set of data, we must specify the evidence that might convince our readers of the validity of a larger and wider application.[7]

Scepticism freed late sixteenth and seventeenth century writers to consider the possibility of a human science, most particularly a science of government: although Bacon illustrated his methodology mainly by the consideration of various kinds of physical phenomena, he clearly stated that he intended the same kind of approach to apply to studies such as history, literature and philosophy. Tuck (1993) has recently looked at the emergence of a modern vocabulary of political theory in writings of this period. Much of this work was primarily theoretical or conceptual but a number of authors, particularly Montaigne, began to reassert the Greek model of systematic inquiry into the varieties of human experience or the narrative of human events as a possible way of investigating the nature of human nature and society. Accounts of human institutions, for example, might be collected to try to induce

[6] See section 4.1.1.4; [7] See section 4.6.1.

what features they had in common and indicate some underlying law or principles of nature which they represented. Although Burke (1981), in a useful short account of Montaigne's work, insisted that it would be wrong to regard him as a proto-anthropologist, he noted that Montaigne's journals show him going beyond literary materials or the reports of others as sources: on his travels in Europe, Montaigne questioned local priests about their theology, visited cathedrals and synagogues and observed a circumcision, an exorcism and a flagellant procession. Throughout he maintained a sceptic's detachment: *Chacun appelle barbarie ce qui n'est pas de son usage*[8] remains one of his most famous aphorisms. Indeed his consistent emphasis on the common humanity of all people, in the notion, for example, that the French were no less strange than, say, the Brazilian tribal peoples that he read about, remains one of his most modern features. Nevertheless, he was writing mainly as a moralist, seeking to instruct his own compatriots about proper behaviour, rather than seeking, as a modern anthropologist might, to understand the thoughts and ideas of the people he observed or read about in their own terms.

2.2.2 Hyperbolic doubt

In the hands of the next generation, Bacon's injunctions were applied in a far more fundamentalist manner. While late sixteenth century writers had generally treated scepticism as a systematic **suspension** of judgement on questions of truth and falsity in order to determine them by rational empirical inquiry, Marin Mersenne (1588–1648), René Descartes (1596–1650), Pierre Gassendi (1592–1655) and Thomas Hobbes (1588–1679) took a more radical, post-sceptical, view that has come to be known as **hyperbolic doubt** (Tuck, 1993:284–94).[9]

Bacon had left two important questions unresolved. The first lay in the speed with which he glossed over the gap between observations and the observed world. He noted the "dulness, incompetency and deceptions of the senses"; the attachment of observers to their prior theoretical schemes; their love of arguments that were really arguments about the application of language rather than about the world itself; and their deference to ancient or fashionable theoretical systems. He called these, respectively, the Idols of the Tribe, the Cave, the Marketplace and the Theatre (Bacon, 1858:54–69). Bacon seems to have seen all of these as corrigible errors. Human beings might be imperfect observers of the world but they could get better. There was no reason, in principle, why nature's secrets should not be unlocked in the way that he proposed. The second question lay in Bacon's vagueness about how induction should be performed. He had said that it was not enumerative and that it must somehow involve methodical comparison but little more than that. His successors attacked both of these questions.

The first is the core of hyperbolic doubt. Basically, it asked how we knew that our observations were an accurate reflection of the external world which we were observing. Why should they be exempt from the solvent of scepticism? Indeed, how could we know that either the observer or the world actually existed? These issues have troubled both philosophers and social scientists recurrently ever since. Indeed, they occur from time to time as the concern of most theoretical scientists, though the most familiar contemporary presentations are probably the debates around social constructionism and post-modernism in social science and the humanities.[10]

In the seventeenth century, they arose from questions about optics, which strayed into what we would now classify as psychology. The study of optics casts doubt on the veridicality of perceptions. If the eye could be fooled by tricks with lenses and mirrors, how did we know what truth was? When we saw an optical illusion, we believed it to be true, just as we believed the evidence of our unaided eyes when we looked at a tree on a clear day. If the illusion were explained to us, we would know that our perception was false. But might it be the case that our unaided eyes had played a trick on us whose secret we had yet to discover? Was there a world out there at all or was it all an optical illusion?

This practical problem has recurred in various forms for generations.[11] Think, for instance, of the belief in Martian canals that persisted until the 1920s or the debate over the extent to which the structures revealed by electron microscopy are artefacts of the electron beam rather than properties of the object. This remains a contemporary problem in areas familiar to HTA. Most health technologies are capable of producing artefactual data, leaving observers with the problem of sorting

[8] Everyone calls barbarous whatever is not customary with him; [9] Arguably, Montaigne's scepticism was more fundamental than Bacon's and led more directly to the position of hyperbolic doubt; [10] See section 3.1.2; [11] See section 3.2.1.1.

out what is a 'real' depiction of the patient or the biological entity being investigated. Although the philosophers we are discussing were concerned with this as a practical issue, their main objective was to see how they could close the gap in a logical way.

What they did, in effect, was to propose that our observations were **signs** of the world. In other words they had the same relationship to the world as a sign does to the object it represents. The word 'cat' is a sign for a small, furry mammal: the use of the sign allows us to discuss members of that species while removed from them in time and space, to talk about them generically or, by adding modifiers, to specify a particular animal – **Jane's black** cat. A sign is not the same as an event, nor is it a representation of an event in the way that a photograph might be said to represent a scene. The world cannot be read off from the array of signs that describe it, though neither can they be wholly detached from it. More shortly, the post-sceptics proposed that we could trust our perceptions: most famously, in Descartes' formula, *cogito, ergo sum*.[12] A thinking body must, at the very least, be certain of its own existence and of the correctness of its immediate perceptions.[13] These perceptions more or less correspond to an external reality because God would not play tricks on his creatures. The reliance on God at the last stage of the argument was quickly seen as a weakness, albeit a dangerous one to dwell on under seventeenth century conditions. Hobbes, for example, who took a number of these ideas into his reflections on how social order might be constructed from original chaos, was persistently threatened for his alleged atheism, precisely because he did not follow Descartes in this move.

However, the application of hyperbolic doubt by the post-sceptics had created a rather different conception of science. Rather than thinking we could understand the world by inferring the rules that governed it, the object of investigation might be our perceptions and our tools for understanding. In effect, perhaps the world was, in a more practical sense also, the creation of our observations rather than our observations being a copy of it. We would need to think much more carefully about the tools that we used.

This concern brought to the fore the other weakness in Bacon's position. His emphasis on induction was designed to mark his break from the dominance of deductive reasoning that had characterised the medieval period. In a modern formulation, induction is defined as 'where we reason from a piece of information, however complex or elaborate this may be, to a conclusion which is logically independent of it.' (Dilman, 1973:29). By contrast, deduction is where 'the relation between premise and conclusion, by virtue of which I am justified in inferring the latter from the former, is internal and can be gathered from the premise and conclusion alone...what the conclusion states is already contained in the premise or premises' (Dilman, 1973:29). Collingwood (1946:52–6), for example, described how medieval historians treated history as an illustration of their revelation of the Divine Plan: events could not contradict God's purpose but interpretation must be stretched to show how even apparently oppositional occurrences in fact contribute to its realisation. The revelation contained within it certain knowledge of the future because its premises determined its conclusions. Historians believed that they could predict what was to come and became relatively uninterested in the detail of human action as subordinate to the Divine Law. The challenge was to make the correct deductions from Divine Law to reveal the story behind some human action rather than to induce that story from the actions themselves. As Collingwood commented, the consequences offer a cautionary tale to more recent attempts to write history, or to do any other kind of social science, on the assumption that it is subject to a law that has no connection to the actual purposes of real human beings.[14]

Bacon's induction, however, was a rather blunt instrument. It retained, in a *sotto voce* form, many deductive elements: Bacon effectively acknowledged that proposed axioms could drive inquiry rather than simply emerge from it. It might also be arguable that a more explicit and observer-independent justification for conclusions as the result of a defensible process of reasoning from observation could help to restore some of the confidence that had been lost in the acceptance of the gap between observation and observed. If we could not know, in an ultimate sense, that some observations were true and others false, perhaps we could know that some observations were based on correct reasoning and therefore to be preferred to others that were not so well justified. If the

[12] I think, therefore I exist; [13] Of course, strictly, *cogito* only demonstrates the existence of the thought, not the thinker; [14] See section 3.2.

world were our creation, then we might try to ensure that we did this with reliable tools.

2.2.3 Immaterialism

The problems of hyperbolic doubt continued to occupy the attention of philosophers for 60 or 70 years, reaching their clearest expression in the work of Bishop George Berkeley (1685–1753). In some ways, Berkeley could be seen as the first radical social constructionist.[15] His position was entertainingly caricatured in a limerick by the twentieth century theologian, Monsignor Ronald Knox:

> *There was a young man who said 'God*
> *Must think it exceedingly odd*
> *If he finds that this tree*
> *Continues to be*
> *When there's no-one about in the Quad'*

Knox was referring to what is known as Berkeley's **immaterialism**. Essentially, Berkeley argued that the idea of a separate world made up of something called 'matter' was unnecessary and irrelevant: what was significant was the way in which human beings organised and classified it, which was the foundation of their actions: "It is evident to any one who takes a survey of the objects of human knowledge, that they are either **ideas** actually imprinted on the senses; or else such as are perceived by attending to the passions and operations of the mind; or lastly **ideas** formed by help of memory and imagination" (Berkeley 1962:65 original emphasis).

What we know is the way in which we construct the world, which does not bear any necessary relationship to it. There is nothing knowable other than minds and their contents.

Put like this, the force of Knox's limerick or Dr Johnson's[16] equally celebrated refutation of Berkeley by kicking a stone should be apparent. Of course it seems absurd to suppose that there is no material world when experience seems to confirm it at every moment: as I type this text, I can feel the hard texture and resistance of the keyboard to my fingers. However, Berkeley's position provided the foundation for much modern qualitative work, even if it is rarely directly addressed (but see Bloor, 1976[17]). Like modern writers, he insisted on the importance of starting from a common sense understanding of the world.[18] What he means by this is that he is not going to introduce any entities like matter in the abstract sense of some of his predecessors, which would not be recognised by ordinary people, and that any technical philosophical statement can be exactly paralleled by an ordinary belief. Berkeley acknowledged four objections to this argument.

- *Everything solid and real is treated in the same way as dreams, illusions and fantasies – we have no way to distinguish between these because everything is in our minds.*
  Berkeley argued that these ideas have a different nature, relation and order which allows us to distinguish between them. The ideas of reality are stronger, are independent of our will and demonstrate regularity and coherence.
- *There is a difference between real fire and imagined fire, between the imagination of being burnt and the experience of being burnt, for example.*
  The different character of these ideas establishes procedures for deciding what will count as real: we can distinguish between imagined burning and real burning by the nature of the sensation and our inability to control it.
- *We can see objects at a distance which we cannot hold in our mind.*
  Berkeley pointed out that we do not have to imagine an external space to judge distance: relative distance can also be exhibited by objects in the imagination.
- *If objects only exist in our perception, does this not mean that when we are not attending to them, they cease to exist?*
  One answer to this is also caricatured by an anonymous reply to Monsignor Knox:
  Dear Sir: Your astonishment's odd;
  I am always about in the Quad.
  And that's why the tree
  Will continue to be,
  Since observed by
  Yours faithfully
  GOD
  In fact, Berkeley made very little use of this argument, that objects continue to exist because they are constantly observed by God. His response was that objects exist as potential triggers of perception but that they only become significant for us as a result of that perception. A table exists for us as a table because we treat a particular physical object according to the properties and usages of a table. When we have a picnic and spread a cloth over a tree-stump,

[15] See section 3.1.2; [16] Samuel Johnson (1709–84) was a distinguished eighteenth century wit, author, scholar and literary critic; [17] See appendix 1 for details of this study; [18] See section 3.2.2.1.

> that stump becomes a table for the purposes of our use for that period of time.

A final clarification might lie in Berkeley's discussion of Newton's work. As he pointed out, the fact that objects fall to the ground, that tides are affected by the moon, that the moon circles the earth and that the planets rotate around the sun were all known before Newton's time. Newton's achievement was to show that they could all be seen as special cases of a regularity called gravity. Some of Newton's contemporaries read this as the discovery of a force:

> But as an efficient cause the force of gravitation would be a mere something we knew not what; to say that apples accelerate towards the earth at a rate of thirty two feet per second because there is a force that accelerates them at that rate is empty talk, as it would be to say that petrol in an internal combustion engine ignites because it has a power of combustion. What Newton did to explain these and related phenomena was to show that they were all cases of the same few basic principles. Gravitational attraction is a concept which has explanatory power because it is a shorthand way of referring to the common feature exhibited by all these phenomena, not because it names an efficient cause of them. (Dunn *et al,* 1992:143.)

At one level Berkeley gave the sort of account of science that a nineteenth century positivist like Mach would offer. The mathematical models of mechanics or physics and the universal laws incorporated into them give us clear and useful tools for describing the relationships between our observations. However, they are also, in another sense, arbitrary descriptions. He used the example of Ptolemaic astronomy, which described the movement of heavenly bodies in terms of a geocentric system of circles along which these objects moved. This proved to be adequate for many human purposes until about the time of Newton when users of the heliocentric model developed by Copernicus, Brahe and Kepler managed to account more accurately for the movements of the same bodies. Which was true? Berkeley said that this is not a sensible question: they are simply alternative frameworks for computation and should be judged in terms of their ability to answer our questions:

> It is one thing to arrive at general laws of nature from a contemplation of the phenomena; and another to frame an hypothesis and from thence to deduce the phenomena. Those who (adopt the Ptolemaic view and by this) explain the motions and appearances of the planets, may not therefore be thought to have discovered principles true in fact and nature. And, albeit we may from the premises infer a conclusion, it will not follow that we can argue reciprocally, and from the conclusion infer the premises. (Berkeley, 1901:230)

Just because we have an answer that works, this does not prove that we have achieved it by the right means. We can deduce a correct prediction from more than one set of premises so the correctness of the prediction does not necessarily mean that the premises we used were correct. This position was quite important in, for example, leading Berkeley to question Newton's distinction between absolute and relative time and space in a way that anticipated the critiques of Clerk Maxwell, Mach and Einstein: the distinction derives from different frames of reference rather than from the meaningless concept of absolute time or space.

Berkeley does not seem to have extended these discussions to the study of society, though he did write on economic issues.[19] However, his importance to qualitative research lies precisely in his immaterialism. In effect, he provides the ground on which many qualitative scholars, from different academic starting points, argue against the proposition that there is anything more to study than the social construction of the world, including the social construction of the world by science.[20] This does not mean that science is a worthless enterprise. As Berkeley acknowledged, the formulation of models and hypotheses that were more predictive than their predecessors enabled much more powerful actions on the world. But the success of a new scientific theory was to be judged by its predictive capacities and it was always vulnerable to alternative descriptions that might lead to more accurate results. We should not make the mistake of assuming that science offered truth. At best it gave us good approximations. In the same way Berkeley's arguments could be turned against social scientific dogmas and against the occult entities that were linked to them. When we talk about social class, for instance, we are really doing the same sort of thing as Newton in describing gravitational attraction and are vulnerable to the same sort of misreading. Social class is a shorthand way of referring to a common feature exhibited by a set of phenomena, not the cause of them. The regularities that we

[19] This work is referred to by secondary sources but does not seem to have attracted much scholarly interest and is inaccessible without considerable primary research which did not seem justifiable for the purposes of this review;
[20] See section 3.1.2.

find in social life are the product of our methods for examining the phenomena for some particular purpose; they are not real in any greater sense than a table is real. A table is a possibility for description, just as a social class is a possibility for labelling some set of observations.

This is not an anti-scientific position. It defines a realm of operation for the natural scientist in a classic positivist sense. Scientists do what they do with more or less success. However, we should be sceptical about some of the claims that are made and recognise the importance of a different kind of inquiry into the constitution of the regularities which they observe and into the uses to which those observations are put. This inquiry might itself be carried out in a rather similar fashion, though later writers were to question whether the specific techniques of natural science could be directly adopted to the study of society or whether it was the underlying philosophical rationale that was important and that this might be operationalised in a variety of ways depending upon the subject matter.

Berkeley's concern with defending the case for immaterialism and locating the role of God within his scheme, which has not been discussed here, meant that questions of method received somewhat less discussion. Berkeley raised the question of what it might mean to say that A caused B once one accepted his view of the primacy of ideas. In brief, his response was that causation was a mental construct from observations of succession. If B always follows A, then our minds formulate a linkage between them. The question of causation, however, became a central issue for Berkeley's near contemporary, David Hume (1711–76), whose legacy, through his influence on the Scottish Enlightenment, was unequivocally foundational to the modern social sciences.

2.2.4 The Scottish Enlightenment

Hume's starting point is very close to Berkeley's conclusions. In his *Treatise on Human Nature* (first published in 1739–40), he declared his programme:

> 'Tis evident, that all the sciences have a relation, greater or less, to human nature; and that however wide any of them may seem to run from it, they still return back by one passage or another. Even **Mathematics, Natural Philosophy and Natural Religion,** are in some measure dependent on the science of MAN; since they also lie under the cognisance of men and are judged of by their powers and faculties. 'Tis impossible to tell what changes and improvements we might make in these sciences were we thoroughly acquainted with the extent and force of human understanding, and cou'd explain the nature of the ideas we employ, and of the operations we perform in our reasonings ... There is no question of importance, whose decision is not compris'd in the science of man; which can be decided with any certainty before we become acquainted with that science. In pretending[21] therefore to explain the principles of human nature, we in effect propose a compleat system of the sciences, built on a foundation almost entirely new, and the only one upon which they can stand with any security. (Hume, 1978:xv–xvi.)

Before Hume's work most philosophers had thought that the route out of scepticism or hyperbolic doubt lay in the discovery of some ultimate correction of our reasoning, some final justification that would define a rational process of knowing. Hume rejected this line, arguing instead that the philosopher's goal must be to understand how we **do** think and to place on one side questions about how we **ought** to think. If we "cou'd explain the nature of the ideas we employ, and of the operations we perform", then we might be better able to judge the conclusions that we had reached and the degree of confidence that we might place in them. The sorts of questions that he asks, then, are empirical ones of the kind that a social scientist might as well ask as a philosopher:

> ...why do we think in the way that we do? Why do we think of some events in the world as causing others, and why do we form the beliefs that we do about things that we have not observed...Why do we think of ourselves as observers of a world of bodies which exist independently of us and which continue to exist when we are not observing them? Why do we think of ourselves as individuals who persist through time...? (Cockburn and Bourne, 1983:14.)

The questions are particularly difficult to answer because of the methodological constraints under which moral philosophy, for which we may read social science, works compared with natural philosophy, for which we may read natural science:

> Moral philosophy has, indeed, this peculiar disadvantage which is not found in natural, that in collecting its experiments, it cannot make them purposely with premeditation and after such a manner as to satisfy itself concerning every particular difficulty which may arise. When I am at a loss to know the

[21] 'Pretending' is used here in the eighteenth century sense of 'claiming' as in the 'Old Pretender', Charles Edward Stuart, claimant to the throne.

> effects of one body upon another in any situation, I need only put them in that situation, and observe what results from it. But should I endeavour to clear up after the same manner any doubt in moral philosophy, by placing myself in the same case with that I consider, 'tis evident this reflection and premeditation would so disturb the operation of my natural principles as must render it impossible to form any just conclusion from the phænomenon. We must therefore glean up our experiments in this science from a cautious observation of human life, and take them as they appear in the common course of the world, by men's behaviour in company, in affairs, and in their pleasures. Where experiments of this kind are judiciously collected and compared, we may hope to establish on them a science, which will not be inferior in certainty, and will certainly be much superior in utility to any other of human comprehension. (Hume, 1978:xviii–xix.)

Hume distinguished two kinds of statement that people can make. The first concerns a relationship between **ideas**, for which he drew on examples from geometry, algebra and arithmetic. These statements can be shown to be definitionally or demonstrably true, in that the conclusions are specified in the ideas themselves. Thus, to say that the square of the hypotenuse is equal to the square of the two sides is to describe a relationship which is given by the system of Euclidean geometry. It is not justified by reference to any number of observations of real triangles but by a chain of reasoning from the axioms of the system. The geometry of an arch is not falsified by the fact that the ancient civilisations of the New World did not discover it and had to construct their buildings with corbels instead. In human affairs, the example is often used of the statements: 'John is a bachelor' and 'John is unmarried'. Since the definition of a bachelor is an unmarried man, if the first is true, then the second must be true.

However, these kinds of statements are not a matter of great interest for Hume, or for most natural or social scientists. What we are concerned about is how we make statements about what Hume calls **matters of fact** (Hume, 1975:25). As he points out, logically the sentence 'the sun will not rise tomorrow' is as intelligible as the sentence 'the sun will rise tomorrow'. Why then do we believe that the latter will actually be the case? How do we form a belief about the future, or about anything else which we cannot directly observe, like the existence of a world beyond our immediate sensations, of other people, etc. Suppose we observe one medical consultation. Why might we believe that other consultations would be like this one? Suppose we carry out a clinical trial for a new drug on 10,000 patients. Why might we believe that the results would apply to the first patient to be prescribed the drug under a full licence?

Hume shows that this is a problem in understanding causation: if we think that A causes B, then we expect that A will always cause B. **Causation** describes a relationship that has three elements.

- **Contiguity in time and space**
  This does not necessarily mean that our observations are of immediate proximity between events, behaviours or whatever but that a chain of contiguous events can be specified which link one observation to another. A modern example might be pressing a switch and causing a light to go on: the switch may be many kilometres from the light – on a motorway information sign, for instance – but the completion of the circuit makes a flow of electrons possible which results in the near-instantaneous illumination of the bulb or diode.
- **Priority in time in the cause before the effect**
  To be able to say that A caused B, A must precede B, as in the example above.
- **A necessary connection between cause and effect**
  Hume saw this as the most important element (Hume, 1978:74). Events may be contiguous or in a temporal succession without being seen to be connected. If the sun rises at 07:00 and my alarm clock goes off at 07:05, we need not suppose that the sunrise has caused the alarm to sound. A necessary connection implies that A **must** be followed by B.

Where do our ideas about necessary connection come from?

Essentially, Hume argued that we observe regular conjunctions – that A is always followed by B which lead us to expect that A will always be followed by B. However, the notion of necessity about this conjunction lies in the mind of the observer. How did Hume get to this conclusion? First, he showed that the expectation that A will always be followed by B depends upon the assumption that 'the course of nature continues always uniformly the same' (Hume, 1978:89). Strictly, we cannot know that the laws of physics will be the same tomorrow as they are today. Indeed since Hume's time, we have established that there are reasons to think that the laws of physics have not always been the same and may not be the same over the entire universe. Nevertheless, if we make this assumption, then we can treat conjunctions as causal in the way that seems to be required for everyday purposes. But

this is an assumption that cannot be independently justified. Hume showed that it is not a relationship between ideas, whose correctness could be logically justified, or that we can deal with it by ascribing some kind of 'power' to A to cause B, in the way that Berkeley, for example, saw the role of God. Hume pointed out that to talk about such a power beyond the sensations of an object or an event simply raises the same problems as talking about necessary connection: what is this power? How can we identify it other than by observations of constant conjunction? Are we not just multiplying imaginary entities and causing more confusion? (Macnabb, 1966:103–12). Just as we saw in our earlier discussion of Berkeley's analysis of Newton, a shorthand description of a common feature should not be muddled up with a cause. Nevertheless, it can create an impression of certainty:

> ...When many uniform instances appear, and the same object is always followed by the same event; we then begin to entertain the notion of cause and connexion. We then **feel** a new sentiment or impression, to wit a customary connexion in the thought or imagination between one object and its usual attendant. (Hume, 1975:78 original emphasis.)

This might seem to be a fairly trenchant criticism of induction. Our explanations of the relationship between events are justified in terms of whether we feel differently about one compared with another. This is a matter of custom rather than reason, a line which has been taken up by twentieth century writers, particularly Wittgenstein and Schutz, to show that it is not purely a private judgement but a judgement in the context of socially shared notions of what it is reasonable or intelligible in a given social situation to treat as connected. We are constrained to link events in particular ways by the threat of being thought to be deluded if we propose other connections. Hume's response, however, would be to ask what alternative there might be? If deductive reasoning only works under the special closed systems of mathematics – and even these have begun to look more fragile in the twentieth century than they did in Hume's day – then it cannot be treated as a gold standard. Mathematics is a domain where the dominance of deductive reasoning is partly a matter of custom, of the shared practices of the community of mathematicians, and partly a matter of its agenda, the relation between ideas rather than between facts or events. Remember, though, that Hume's ambition is not to **correct** our natural modes of reasoning but to **understand** them. What he is saying is that we are made this way, that this kind of causal explanation is the best that we can achieve and that we would be better employed trying to see what we can do with it rather than supposing we could remedy its apparent failings.

Hume also performed the important task of integrating natural and social sciences in a single methodological framework. As Ayer (Dunn *et al,* 1992:258–9) pointed out, this does cause some problems in reconciling talk about observed regularities and inferred causes in human actions with the actors' own protestations of free will. Near-contemporaries like Smith and Ferguson began to explore the question of how individual actions produced institutional regularities which were, in some sense, independent of the actors' own intentions (Schneider 1967:xxix–xlvii). This is a recurrent issue for all the social sciences: how far is action determined and how far is it choice?[22] There is also the argument that Collingwood, for example, put forward in noting that a Humean approach has difficulty in dealing with social change. The force of its analogy between natural and social science leads to an assumption, which is probably not necessary to the argument, that human nature is uniform and unchanging in the same way as Nature itself:

> Coal tar, for example, once its chemistry is understood, ceases to be refuse and becomes the raw material of dyes, resins and other products, but the fact that these chemical discoveries have been made in no way alters the nature of coal tar or its by-products. Nature stays put and is the same whether we understand it or not. (Collingwood, 1946:84.)

Such an assumption makes a history of human nature impossible because the possibility of history implies that human nature is a product of historical process while Hume would tend to treat it as a presupposition. The uniformity of human nature is assumed rather than demonstrated so that accounts written in this period still tend to have a deductive character, stories told to show the uniformity and unchangingness of human nature rather than to examine the link between human nature and historical context. Against Collingwood, however, we could argue that if some measure of uniformity is not assumed, then other times or places become literally incomprehensible. If human nature is so variable that we cannot infer regularities in the way that Hume implied, then society itself may be impossible.

[22] See section 3.2.2.

Nevertheless, Hume did argue that, just as he read Newton as claiming to avoid any generalisation not based on observation or experiment, so too we might study human beings and, at least by extension, the societies in which they lived by the same methods, saving his comments on the difficulties of experimenting with people. But, if Hume showed that inductive reasoning was endemic in all the sciences, he did not prescribe any particular methods for implementing this. If one looks at the eighteenth century figures who are generally considered to be the first recognisable social scientists – Smith, Ferguson and the other Scots whose names are linked primarily to economics and anthropology; Montesquieu in France for sociology, and Vico, an Italian historian – they employ a very wide range of methods. Smith, for example, used statistics, historical accounts, travellers' reports and his personal observations in *An Inquiry into the Nature and Causes of the Wealth of Nations* (1976). The analysis of the division of labour which opens this book is sustained by the famous example of pin-making, 'I have **seen** a small manufactory of this kind where ten men only were employed...' (Smith, 1976:i,8–9 emphasis added). At other places, of course, he also drew heavily on the tradition of political arithmetic and the various attempts to produce calculations of national wealth and flows of capital or trade that had been developing since the middle of the seventeenth century (Shaw and Miles, 1979).

Smith was not a self-conscious methodologist, though he acknowledged the extent of his debts to Hume. If we look at his methods of argument, though, we can see both classic induction, in the sense of the presentation of a series of instances and the attempt to infer an underlying principle, which can then be connected to other principles in his overall scheme, and something which looks rather like deduction, at least in a loose sense. His discussion of apothecaries' profit, for instance, begins with a statement attributed to common sense, in the form "apothecaries make excessive profits". He then rejected this by pointing to the empirical evidence of the elements of skill and trust that are involved in the work, from which he concluded that a large part of what is called profit is in fact better regarded as the reasonable wages of this labour. The apothecary cannot charge for his advice so he must recover the value of his time through the price charged for his drugs (Smith, 1976:i,125). Today, we would probably call this loosely a **hypothetico-deductive approach**.

However, this is really rather different from the deductive approach associated with modern neo-classical economics, which has far more in common with the approach to mathematics taken by Hume. Unlike Smith's political economy, neo-classical economics is not an empirical discipline but a set of axioms from which various conclusions can be deduced. This system is a system of ideas, where truth is definitional or demonstrable rather than empirically verifiable. Smith was far more concerned with matters of fact and of trying to specify the necessary connections that exist between economic and related events and which, if correctly specified, will enable observers to make predictions about the effects of changes in the world. Suppose there is a bad harvest. Will the price of grain rise? Suppose there is a tax rise. Will this increase or decrease government revenues because people can reduce their consumption or substitute alternative goods or services for the taxed item?[23] As Hume showed, we cannot know the future with the same degree of certainty that we can know that a mathematical proof is correct. However, this is not to say that we cannot try to produce a set of necessary connections that we can expect to work most of the time. Indeed, this is the best that we can hope to achieve and we should not mislead ourselves by thinking we can achieve more certainty. The criteria for demonstrating a connection between ideas are just different from those for demonstrating a connection between observations or events, facts in Hume's sense. But it is not clear that there is a large practical difference between citing a number of observations and describing a principle that could explain them and stating a general principle and then showing it not to be capable of general application by citing a contrary case, thus requiring revision of the statement.[24]

Hume's work had a considerable influence on the Scottish Enlightenment writers whose work laid the foundations for most of the modern social sciences. However, contemporary practice probably bears a stronger imprint from the way in which the same issues were revisited by two other scholars. One was John Stuart Mill (1806–73), particularly in *A System of Logic*, which he began to work on in 1830/31, first published in 1843 and revised through eight editions until 1872. The other was Immanuel Kant (1724–1804), particularly in his *Critique of Pure Reason* (1781) and *Critique of Practical Reason* (1788) (Kant 1964, 1976).

[23] A modern economist would phrase this as a question about the price elasticity of demand for the item;
[24] See section 4.6.1.

2.3 What is a social science?

The positions outlined by Hume and Berkeley provoked a variety of hostile responses. Many people found the doctrine of immaterialism unpalatable, while others, notably Kant, attacked the idea that the human mind was formed entirely by experience. We were, in effect, **hard wired** to perceive the world in certain ways, though Kant did accept that all we could have reliable knowledge of was our perceptions. Hume's ideas on causation also provoked a variety of attempts to show that there was something intrinsically necessary about necessary connections rather than this being wholly in the mind of their observer. Mill was less interested in the epistemological issues than in examining the procedures of inference by which people came to think that they had reliable knowledge about the world. In the process, he worked out a much clearer account of the relationship between inductive and deductive reasoning (Fletcher, 1971; Nagel, 1950). For our purposes, Mill makes three important contributions: his discussion of the ways in which induction might be carried out and evaluated according to certain canons of procedure; his discussion of what we might mean by a causal law; and his application of both of these to the fields of psychology, sociology, economics and history. Before examining these, however, we shall return to the epistemological issues opened up by Berkeley and Hume and examine Kant's response.

2.3.1 Experience, reason and method

Although much of the practice of modern social science in the English-speaking world draws its models from the Scottish tradition as developed by Mill, sociology and, to a much lesser extent, psychology, anthropology and economics, also bear the imprint of the rival developments on mainland Europe. The focal point of these lies in the work of Kant. As Kant himself stated, many of his arguments were elaborated in response to Hume, though he also reacted against the work of the Prussian philosopher Gottfried von Leibniz (1646–1716). Kant's writings are justly considered difficult and obscure and are only given a brief summary derived from secondary sources here. However, they became foundational to much German and some American qualitative social science and some acquaintance with their principal arguments seems unavoidable, whether to understand adherents or critics.

Leibniz had argued that all knowledge was founded upon a process of reasoning that could give access to an objective, observer-independent view of the world. We were all born with certain innate principles of understanding, which constituted the axioms out of which knowledge of the world could be built. The world itself was full of objective entities which he called **monads**. We might see these from different points of view but those different perspectives did not change the inherent nature of the object or of the relationships between the objects. The relationships which we perceive between objects are necessary ones: we are constrained to perceive them in that way because that is what they really are. The contrasts with Hume are evident. Hume rejected the possibility of innate ideas or reason. Our knowledge of the world was founded on our experience of it. Knowledge of an objective world is simply a claim to a degree of consistency and constancy in our impressions of it which creates a sense of its independence. The relationships between perceived objects are also the result of experience and of an expectation that A will be followed by B: this is not a necessary conclusion.

Kant rejected Hume's position as fundamentally destructive to the possibility of science. Causation was not an arbitrary matter. At the same time, he also rejected much of Leibniz's position on the innateness of reason. His argument began from an attempt to reconcile these two conclusions. If Hume was wrong to argue that experience alone was able to provide a basis for knowledge, then Leibniz was also wrong to argue that reason alone was sufficient. Experience provided the contents of our knowledge, but reason was also necessary to provide a form or structure. In a sense this prefigures contemporary discussions about the theory-driven nature of observations.[25] We cannot know what to notice about the world unless we have a prior theory which indicates where to look: at the same time our theory should also be subject to challenge if the observations are not as anticipated. The result, Kant proposed, is a form of knowledge that is genuine, objective and not observer-dependent. It is impossible to have knowledge of the world as it inherently is, in the manner that Leibniz had suggested through the exercise of pure reason. We always know the world from our own point of view. However, the material nature of the world constrains the possible points of view. Objects do not depend upon my perception for their existence, as Berkeley and to an extent Hume had implied. It is their perceivability that is the mark of their existence. Experience has an objective

[25] See section 4.6.1.

referent, even if that referent can only be known through perception.

Kant called this doctrine **transcendent idealism**, for the way in which it overcomes the notion that knowledge is purely a set of ideas of the world. Part of our task is to focus on the presuppositions that make experience possible. What are the *a priori* features of our minds that mean that we have the ability to know the world around us? Kant insisted that this is not the basis of an empirical psychology but is a theory of cognition. A being that is able to relate to its surroundings in a particular way must have a particular kind of cognitive organisation to make this possible. Here Kant was retaining elements from Leibniz about the preformed nature of our ability to reason. However, he went on to emphasise the way in which we come to know things in a more dualistic fashion. On the one hand, we draw knowledge from our sensations of the world, as Hume and the empiricists supposed. On the other hand, we have a knowledge of concepts that we apply in organising experience. The former is passive, while the latter is active. In Scruton's (1982:25) words: "A mind without concepts would have no capacity to think; equally a mind armed with concepts but with no sensory data to which they could be applied would have nothing to think **about**" (original emphasis).

Judgement is a synthesis of these two processes (or structures – Kant does not seem to be very clear about which). By concepts, Kant meant the fundamental categories of organising knowledge of which the most important are the ideas of substance and of cause. A substance is that which can exist independently of our observations and which is capable of supporting the properties of those observations which appear to depend upon it. The world must seem to conform to these categories for us to have knowledge of it. This is Kant's reply to Hume: experience is already structured before we perceive it and we are obliged to perceive it in ways that correspond to this structuring.

However, that structure is not fixed and unchanging in the way that Leibniz had proposed. It is a world of ordinary objects and processes in time and space. To quote Scruton (1982:31) again:

> The world consists rather of ordinary spatio-temporal objects. The philosophical proof of objectivity establishes the existence, not of an abstract perspectiveless world but of the commonsense world of science and everyday perception: the very world which both Humean scepticism and Leibnizian metaphysics had thrown in doubt.

This knowledge is further grounded in the way that we think about ourselves, which Kant calls the **transcendental unity of apperception**. By this he means our sense of ownership of our experience – apperception refers to the experience of a perception as definitively **mine**. Scruton referred to the example of Mrs Gradgrind in Dickens's *Hard Times* knowing on her death bed that there was a pain in the room but not that it was hers. It is 'transcendental' because it is again a fundamental that cannot be derived from experience. I have to be able to experience my perceptions as mine in order to separate myself from the world around me. At the same time, I have to be an object of the same kind as other objects to be able to operate in this world. Again, this requires a world which has certain properties – a sense of continuity, of identity through time and space and to have a sequential ordering such as to be able to distinguish past, present and future.

In contrast to the empiricist tenor of British developments, then, Kant laid out a position where theory and data were given equal status: indeed in certain respects, theory could be prior to, and incorrigible by, data. As this chapter develops we shall see a range of consequences from this. However, two are of sufficient general importance to bring out here. The first is Kant's distinction between **noumenon** and **phenomenon**. A phenomenon is an object of possible experience, something which happens or which is there in the world and available for us to react to or to act upon. Later writers would explore the topic of phenomenology, what it took for us to perceive a possible experience and to act in relation to it. Noumenon refers to what might be called a mental construct, something which forms part of a system of ideas or thought and which cannot be considered as an object of experience. In some respects this resembles Hume's discussion of the difference between matters of fact and relationships between ideas. The appropriate analysis of these two different kinds of object is likely to be different. Phenomena are examined in relation to the world of objects in which they occur while noumena are examined in the context of a system of ideas which has no necessary empirical foundation. The latter is the seductive world of pure reason, which draws us towards the illusion of perspective-free, objective knowledge. While this is useful as a regulative ideal, in directing us to a search for orderly thinking and for more general explanation of specific phenomena that advance our understanding of the empirical world, pure reason is ultimately a self-contained system of speculation which has little other value.

Kant's ideal is what he calls **practical reason**. This unites judgements of fact and value. As rational beings we recognise the difference between knowing that something is true and knowing what to do as a consequence. However, Kant effectively dissolved the difference in arguing that to know something is in itself a moral act that requires a consequence. The exercise of practical reason is not simply a technical question of the selection of means towards an end, in the way that Hume and the other Scots had tended to suppose. In their scheme, the ends of action had derived from our passions, the unregulated aspects of our nature.[26] Against this, Kant proposed that our goals could also be an object of reasoning and could not be separated from our deliberation over means. If moral action were grounded in reason, then it could become objective, rather than subjective as the Scots implied. The moral schemes of practical reason bind everyone unconditionally.

The importance of this line of argument has been in enshrining in much subsequent social theory the notion of moral critique, though Kant himself had little to say on the subject of social as opposed to natural science. His position, however, leads to the argument that social theory can define objectively what constitutes a moral action or a moral life and social theorists, in particular, can supersede Goethe's poets as the legislators for humankind. The fact/value issue and the implications for politics and moral judgement has been an important dimension of debates in German sociology for the last hundred years or so (Adorno *et al,* 1976). This particular aspect will not be examined in detail here, except to note the origins of the concept of critical theory and of the claim that social scientists have a duty to be social critics rather than social technicians.[27]

The immediate reactions to Kant's arguments were generally hostile, with a generation of distinguished German philosophers responding with neo-Humean challenges to his compromise (Beiser, 1987). They were simply unconvinced by his arguments against the primacy of experience, by the presupposition of his fundamental categories and by the problematic relationship between concepts (noumena) and experience (phenomena). If they were so separate, how could one structure the other? How did we decide that a phenomenon stood in a particular relationship to a noumenon? This last remains a fundamental problem for all social science.[28] A more sympathetic response came rather later in the work of Schiller and Hegel (Fletcher, 1971; Hawthorn, 1976). Hegel, in particular, argued that the study of history was capable of demonstrating the necessarily rational character of the world, which we were obliged to respect and reflect in our subjective understanding. In effect, he tacked back towards Leibniz's position, though without accepting the doctrine of monads. For Hegel, knowledge was not a matter of experience: rather experience carried an inherent structure of reason that also embraced us as knowers and actors in that world. We were rational beings because rationality was a condition of our existence in the material world that we inhabited. Rationality represented the centrality of ideas or spirit in the pre-constitution of experience. Where Kant had stressed the interaction between theory and practice or ideas and objects, Hegel argued that ideas or theory were necessarily prior. This drew him much more towards social science, or at least some notion of history or jurisprudence, which would document the working out of spirit in the specific historical forms of states and civil society. Marx modified this position in reasserting the primacy of the material world and of experience and cautioning against the partial nature of most theoretical understandings that were not derived from these sources.

It was only in the 1870s, however, that these epistemological debates were explicitly linked to the methodological practice of the social sciences. Although Marx had undertaken a number of empirical inquiries, which included both qualitative and historical studies associated with his journalism and his polemical writing and attempts to collect information by means of questionnaires, these do not seem to have been explicitly linked to his theory of knowledge. With the greater professionalisation of German social science, these questions became unavoidable. How did we justify our claims to know something? This is reflected in what has become known as the *Methodenstreit,* the quarrel over methods, which first opened between two economists, Schmoller and Menger and has divided most German social scientists for the last hundred years (Adorno *et al,* 1976:ix–xliv).[29] Summarily, it has been a debate over issues, which should by now be familiar: are the social sciences exact or historical; deductive or inductive; abstract or empirical? Menger argued that the world of phenomena could supply two types of knowledge:

[26] This is somewhat less true of Smith who explored the way in which passions could be constrained by interests, which could be grounded in reason (Hirschman, 1977); [27] See section 3.1.3; [28] See section 5.2; [29] See sections 3.2 and 5.2.

a knowledge of individual concrete instances and a knowledge of forms or types and the relations between them. In some respects, this resembles the Kantian distinction between phenomenon and noumenon, except that Menger does not separate types from empirical occurrences in the way that Kant separates noumena from the material world. Types are displayed in their individual manifestations, which, in turn, provide a basis for their characterisation and theorisation. It is our understanding of the typical that makes prediction and control of events possible.

Menger distinguished three types of economic science: historical economics, which concentrates on the description of individual cases; theoretical economics, which is the analysis of types; and practical economics, which is the techniques of economic management. Historical and theoretical economics were quite different enterprises. Historical understanding involved the investigation of individual cases and their unique history, while theoretical understanding involved the recognition of individual cases as instances of types. The latter led on to the specification of models that linked types through laws that determined their relationships. Schmoller challenged the separation between history and theory, arguing that it tended to lead to an obsession with abstract models devoid of adequate empirical referents. Historical studies could also generate rules and hypotheses. This involved a rather different understanding of the subject matter of economics: it was not the relationship between quantities of goods or services but the study of economic actions. Those actions could only be analysed with reference to the knowledge and intentions of the actors and not reduced to mathematical models.

One route out of this controversy was offered by the neo-Kantians, Windelband and Rickert, writing in the 1890s, with their distinction between *Naturwissenschaften* and *Geisteswissenschaften,* natural sciences and historical, cultural or moral sciences. The former are nomothetic, ordering disciplines concerned with the specification of laws and the formulation of generalisations, while the latter are idiographic, concerned with the description and understanding of specific individual occurrences. However, and this point is often overlooked, the distinction does not necessarily refer to the object of study so much as the purpose for which the study is carried out (Freund, 1968; Adorno *et al,* 1976). In other words, one might have both nomothetic and idiographic studies of the same subject matter for rather different purposes. Thus a natural science of society might be concerned with the discovery of laws or law-like regularities relating to typical forms of individual or institutional action, while a cultural science of society is focussed more on issues of value. What is it about a society that leads us to think about that society in a particular way?

Cultural science is defined by its attempt to understand the spirit *(Geist)* of a particular social or historical institution. This was a context-specific sense of just what was unique about that time or place, as against the generalising thrust of natural sciences. How could this sense be derived? Dilthey, a philosopher with interests in both history and psychology, was the first to set out the notion of *verstehen,* a systematic understanding of the inner consciousness of the analyst derived from their own experience and from a sympathetic reconstruction of the experience of those being studied (Hughes, 1959). There are important affinities here with Smith's position in *Theory of Moral Sentiments,* though it is not clear whether Dilthey was directly influenced by Smith. However, Dilthey's ideas, and indeed the Neo-Kantians generally, certainly influenced the American pragmatist philosophers James and Mead, and both sets of ideas were circulating in the formative years of Chicago sociology[30] and in the definition of the programme that came to be known as symbolic interactionism (Rock, 1979; Joas, 1985; Dingwall, 1997b). The problem that Dilthey never fully solved, and that many others have grappled with, is how to specify what was meant by a process of sympathetic experience and exactly what procedures were followed by the investigator in achieving this kind of understanding in such a way that the claims could be adequately evaluated.[31]

This intellectual separation led to institutional separation. Sociology, economics, politics and psychology all divided between positive and normative schools. The former saw themselves primarily as scientists, conducting a value-free inquiry into the way in which a particular domain of human interest worked. The latter saw themselves more as critics, raising questions about how that domain **ought** to work. For sociology, the object of study was most clearly stated by Weber as "the scientific investigation of the general cultural significance of the socio-economic structure of human community" (Weber, 1963:373). Values and, by implication, culture were a proper object of inquiry. However,

[30] See section 2.4.3; [31] See section 5.2.

a scientific study could not address their validity. In the strictest sense, values had an irrational or, perhaps better, a non-rational foundation. Sociology could point to discrepancies between professed values and actions said to derive from them. Suppose, for example, that a health service manager claimed that she was morally committed to equal opportunities but then used recruitment strategies that systematically failed to search for candidates among certain population groups. In a Weberian approach, we could not justify her original commitment but we could show that her actions were not rationally linked to it. We could, however, also seek to understand her original commitment through the method of *verstehen*. In effect, Weber took this out of the exclusive context of the cultural sciences and argued that it could also be used in a positive scientific manner (Freund, 1968:39). He acknowledged that some kind of intuitive element was unavoidable unless social science was to be purely a matter of counting behaviours (Hughes, 1959). We could count all the occasions on which people scratched their nose in an auction room but the results would be entirely trivial unless we knew why people had done that – to relieve an itch, to dislodge a fly or to signify a bid. But this intuition had to be disciplined by attention to the details of context and the extent to which they corroborated the intuitive conclusions and by the consistency of intuitive conclusions on one occasion with those on others, as formulated in more abstract and typical fashion. Are the conclusions from *verstehen* on one occasion consistent with those that we would, in general, expect to find on others? If not, could we account for the differences in ways that would be logically relatable to our previous findings?

Weber's conclusions on method can stand as representative of the most influential outcomes from this debate. First, he rejected the idea that the only valid way of doing science was by quantification and measurement. These are means towards an end rather than ends in themselves.[32] If the goal of scientific inquiry is to produce a disinterested statement about the condition of a world which, if not exactly observer-independent, has a substantial degree of autonomy and conditioning of the possibilities of observation, then any appropriately disciplined procedure can be used. It happens that mathematics was the first such procedure to develop but it is neither an exemplary nor a universal model. Mathematical concepts are one way of organising the world but, like all concepts, they select from the infinite variability of reality. The selective and finite nature of concepts means that they can never aggregate to the sum total of reality. Nevertheless, coherent and logically organised systems of concepts can produce a firm enough account of a particular domain to be a basis for effective actions (Freund, 1968; Aron, 1970). Quantification is one way that such systems can in some circumstances be produced and Weber himself made extensive use of it, especially to inquire into economic issues, many of which are, of course, preconstructed in numerical forms as a result of the workings of the price mechanism.

Second, however, Weber, as noted above, rejected the idea of pure intuition from experience or empathy with another. Intuition is an emotional process, which forms part of an aesthetic rather than a scientific understanding of the world. Experience and intuition can only be part of a scientific enterprise if they are transformed into concepts and subjected to appropriate verification procedures. Nevertheless, they are essential aspects of a sociology which examines the construction of society from the everyday actions of ordinary men and women and the relationships that they constitute (or an economics which derives its models from data on individual transactions that are aggregated into larger models, or a psychology which develops models of cognition from observations of individual cognitive acts, etc.). These specific observations may be assembled into ideal types, which select and abstract what appear to be key properties and which may then be talked about at a level of generalisation that is detached from the original data. Both general laws and the examination of specific cases have a role in a Weberian sociology.[33] Although Weber used observational examples in setting out this methodological position, most notably that of the problem of understanding the motivations of a man about to cut down a tree, his own empirical work was predominantly historical in nature. His study of the origins of capitalism in Europe, for example, proceeded by a systematic examination of both the material and the cultural conditions and by comparison with cases where the material conditions were present but the cultural ones were absent (China) and where the cultural ones were present and the material absent (India). His initial intuition that Protestant religious beliefs were an important factor in making possible the transformation from a feudal to a capitalist society was formalised and

[32] See section 3.1; [33] See section 4.6.1.

tested through what were, in effect, natural experiments or controls.[34]

In general terms, it is hard to point to a specifically Weberian qualitative sociology, except in terms of historical writing. With some partial exceptions in the field of industrial sociology, most qualitative work has drawn its direct inspiration from other sources. However, Weber's arguments have been very influential in legitimising qualitative research, especially in reference to the role of *verstehen* and in the notion of an interpretative sociology. In practice, though, Weber's caution about *verstehen* has often been overlooked and he has been used to justify practices which come closer to Dilthey's original formulation, which separates the natural and the cultural sciences.[35] Similarly, although Weber writes about an interpretative sociology, his emphasis is very much on the latter part of the term. Interpretative understanding is one means of developing a general science of society which aspires to the same generality and objectivity as any other science. Even in his case studies, Weber's concern is to use unique events to discover regularities or general principles, if not laws in a strict physical sense. The rise of capitalism in Europe may have been a unique event but it was one which arose predictably from the conjunction of specific material and cultural conditions. In the same way, each hurricane that tracks along the Florida seaboard is a unique event; it arises, however, from a well-defined interaction between air pressure systems, currents and water temperatures in the Gulf of Mexico. If this interaction occurs, we can expect a hurricane but we have difficulty in predicting whether it will run up the Gulf Coast and the Mississippi Valley or up the Atlantic Coast and into the Carolinas.

This section has concentrated on the Kantian tradition and left its positivist adversary characterised rather by implication. In returning to the discussion of Mill, we can bring out more sharply some of the things that the Germans were arguing against, though Mill should not be seen as an unreflective representative of positivism.

2.3.2 Induction, deduction and a science of society

The notion of positivism has given social scientists a great deal of trouble over the years, partly because it has often been used as a 'boo-word' to denounce points of view that the user finds uncongenial rather than as a specific description of an intellectual position. As a technical term, it was introduced by the French writer, Auguste Comte (1798–1857) in his *Positive Philosophy* (published between 1830 and 1842). Comte, who was also the person who gave us the word **sociology**, exemplifies the nineteenth century belief in the power of natural science as the paradigm of all valid knowledge and as the means of solving the great practical problems facing humankind. The sciences were organised in a hierarchy, which was founded on those studies that related to phenomena most remote from human control and which gradually moved towards the study of humanity itself. In his scheme, physics would be the foundation of chemistry, which would be the foundation of biology, which would be the foundation of sociology. Comte's position should not be confused with the contemporary position held by some sociologists of science, technology and medicine that sociology is the key to all sciences because they are all social institutions and social constructs.[36] It is better understood as a statement that the laws of each lower tier constrained the possibilities of the ones above it. The possible laws of social organisation, then, were limited by the laws of biological organisation, which were limited by the laws of chemistry, which were limited by the laws of physics. By understanding the laws of social organisation, which would look something like the laws of Newtonian physics, we would be able to design better societies just as the understanding of physics had allowed us to design better machines.

Since Comte's time, the notion of positivism has gone in slightly different directions among philosophers and social scientists. For philosophers, positivism is a position that emphasises phenomenalism, the aspect of experience as the basis of valid knowledge. It rejects Kant's separation of phenomenon and noumenon, regarding the latter as abstract, speculative and with no grounding in the real world of the senses. It also rejects the immaterialism of Berkeley, Hume and Kant: the world that we know is the real world. Our observations have not been passed through any filters before they reach us so we know it directly. This world contains no inherent judgements of value. Truth is a matter of correct description and ideas like justice or beauty have no referent in it.

In the social sciences, as Giddens (1974) noted, positivism has three specific methodological implications.

[34] This work has been overtaken to some extent by further historical evidence and more detailed examination of the cases. However, it remains important as a methodological model for sociologists; [35] See section 3.1.2; [36] See section 3.2.1.

- *The methodological procedures of natural science may be directly adapted to social science.*
  There are two specific consequences from this. The first is the belief that valid knowledge can only be obtained by directly copying the methods of the natural sciences with their emphasis on the quantification of observations, on formalisation and abstraction and on the use, where practicable, of experimental methods to test deductions from laws. The second is the belief that society and social life are objects of study in the same sense as the material world of physics, chemistry and biology. A social class, for example, is an object of the same kind as an amoeba. Positivism discounts the relevance of the subjective understandings and intentions of the actors which the proponents of *verstehen* were trying to recover and incorporate in their schemes of explanation. It should be noted that both of these beliefs rest on a rather nineteenth century model of science. The actual practice of contemporary scientific inquiry seems to be much more varied and physicists, at least, seem to be less certain that they are dealing with an observer-independent world of objects than their predecessors.
- *The outcomes of social scientific inquiry will look like those of natural scientific inquiry.*
  The objective of social science is to develop a set of laws or law-like statements about human beings, their society and its organisation. These laws will look something like the laws of thermodynamics, for example.
- *Social science is a technical enterprise with no necessary value implications.*
  This statement needs to be understood in two ways. In the first sense, it expresses a scepticism about the claim that social scientists can make normative prescriptions about what a good life or a good society would be like based on the authority of their discipline. As we have seen, this was an important strand for some of the nineteenth century writers and, indeed, for Comte himself. However, a positivist social scientist cannot also be a critical theorist, though, of course, he or she may have views as a private citizen which do not have the same rational foundation as the knowledge of science. As an elaboration of this, which is important for the present review, the positivist social scientist is also sceptical about the relationship of his or her work to policy. The objective is to achieve a scientific understanding in its own right which is neutral with respect to the value or policy implications. In principle, a student of organisational behaviour might study the efficiency and effectiveness of a concentration camp as means to an end in exactly the same way as he or she might study a convent.

As Giddens pointed out, individual social scientists do vary in their specific positions on each of these issues and they do not necessarily go together as a package. In particular, there is clearly a major body of policy-oriented social science that adopts the first two assumptions but rejects the third. It is, however, useful to be reminded of the flimsiness of the foundation on which that prescription rests. We may, for example, establish by the methods of social epidemiology, that certain kinds of health risks are unequally distributed in the population: this does not oblige us to treat that as a problem which should be solved in some fashion. The judgement that an unequal distribution of health risk is a problem, which we as a society ought to do something about comes from a different source and does not, in the positivist scheme of things, have the same weight as the empirical finding.

The positivist programme has been very influential in all of the social sciences over the last hundred years. Indeed, at various times, it has enjoyed a near-monopoly of legitimate research, especially in psychology and economics. However, its clarity and focus are both strengths and weaknesses. Its position on the objectivity of the social world, which has echoes of the earlier work of writers like Leibniz, has been difficult to sustain in the face of persistent evidence that the social world is different and that the parallels between the study of nature and the study of human beings and their society cannot be exact. Its determinism, the idea that human actions are governed by laws of behaviour, may be superficially attractive but has found it hard to resist the consistent demonstration that human action may be better understood as a creative act of rule-orienting rather than an automatic act of rule-following. The latter points towards inductive rather than deductive approaches as likely to yield more satisfactory results. Its value-neutrality has often been compromised into furnishing justifications for prior normative positions, which have themselves been excluded from analysis by the split between science and policy. The neo-Kantian programmes, for example, have been much more insistent on the need to accommodate both the study of action and the study of values. These are distinct tasks, which may require different methods but are ultimately interdependent.

We can examine some of these issues in more detail through a discussion of the work of Mill,

who was one of the first admirers of Comte in England but who became somewhat less enthusiastic as he worked through the implications of a positivist position. Mill (1974:833–5) began his discussion of what he still calls the moral sciences by noting that his exposition of the basic principles of scientific method must *a fortiori* apply to them as well. The task is to see how they may best be applied and to what extent their limited success results from inappropriate applications or to lack of skill in the use of appropriate ones. He identified three models that might be used: chemical, geometrical, and physical.

- The **chemical** model takes the use of experiments in chemistry as its exemplar for social sciences. Mill's particular target was the practice of historical and political studies (1974:879–86). He argued that the basic conditions of experimental method simply cannot be met in these sciences: we cannot control all the possible sources of variation, we cannot list all the possibly material observations, the timescale for results is so long that environmental change will probably render them invalid and so on.
- The **geometrical** model takes Euclidean geometry as its exemplar (Mill, 1974:887–94). Its error is to have adopted an overly simplified form of deductive study. Unlike mechanics or astronomy, geometry cannot handle the idea of conflicting or contradictory forces. Geometrical theorems cannot be invalidated by other geometrical principles. Mill discussed this in relation to the classical approach to economics and to certain types of political theory. Starting from the equivalent of a Euclidean axiom, that people are motivated purely by their own self-interest, these writers have been able to construct an elaborate deductive scheme which makes a variety of predictions about human behaviour in specific cases. While Mill (1974:893) admired the intellectual elegance of the results, he was critical of the omission of other determining factors:

  > It is unphilosophical to construct a science out of a few of the agencies by which the phenomena are determined, and leave the rest to the routine of practice or the sagacity of conjecture. We either ought not to pretend to scientific forms, or we ought to study all the determining agencies equally, and endeavour, so far as it can be done, to include all of them within the pale of the science; or else we shall infallibly bestow a disproportionate attention upon those which our theory takes into account, while we misestimate the rest, and probably underrate their importance.

  The result, Mill argued, is that deductive schemes, like that of economics, focus only on those elements of human behaviour that can be accommodated within their models. However, human nature is not governed by simple laws in the way that this approach assumes. It needs to be understood in a context of interacting principles, where a change in any one has implications for the whole system.
- Mill's preference then was that social science, which "by a convenient barbarism, has been termed sociology", should be modelled on **physics** (1974:895). His argument starts from the observation of the successes of the physical sciences in handling the notion that events can be explained as the product of several different laws, which interact in a particular way to achieve a particular effect. We might think of the discovery of gravitational 'dead spots' where the pull of different bodies intersects and cancels out, for example. Social phenomena will be explained by similar laws, though vastly more complex calculations will be needed to deal with the number of interacting bodies and tendencies that are involved. Mill was fairly sceptical about our ability actually to do this and developed what has come to be called the inverse deductive method. This is the forerunner of modern hypothetico-deductive approaches in that Mill suggested that we may proceed by reasoning from laws or principles that we believe to hold true to make predictions about the likely course of events which may then be verified by subsequent observations. Unlike traditional forms of deductive method, however, our confidence in the conclusions does not derive from the initial reasoning but from its verification by observation. However, he also noted that the complexity of social life often means that we end up reversing the process: we make specific observations which we then connect by a process of reasoning to more fundamental laws or principles.[37]

Mill stressed that the social sciences cannot expect to produce conclusive predictions or determinist laws. Social life is much too complicated. At the same time, however, a knowledge of tendencies may still be useful:

> It is not necessary for the wise conduct of the affairs of society, no more than any one's private concerns,

[37] See sections 3.2.1.2 and 4.6.1.

> that we should be able to foresee infallibly the results of what we do. The aim of practical politics is to surround any given society with the greatest number of circumstances of which the tendencies are beneficial, and to remove or counteract, as far as is practicable, those of which the tendencies are injurious. A knowledge of the tendencies only, though without the power of predicting their conjunct result, gives us to a considerable extent this power. (Mill, 1974:898.)

Because all social phenomena are influenced by all other social phenomena, we can never expect the exact conditions of one historical or empirical observation to be reproduced on another occasion and to lead to the exact same results. If social sciences have a deductive character, this is not in the sense of producing the kind of theorems that we would find in geometry, where we can deduce specific conclusions from a limited number of principles, but rather to be able to identify a range of principles that might be particularly involved in some specific situation and to suggest what conclusions may be more likely than others. Mill (1974:913) also noted the problem of feedback in social science, the reaction of effects upon causes: "The circumstances in which mankind are placed, operating according to their own laws and to the laws of human nature form the characters of the human beings; but the human beings, in their turn, mould and shape the circumstances for themselves and for those who come after them".

Mill identified the key evidence for the existence of laws as the discovery of statistical regularities. Much as the French sociologist Durkheim would do about 20 or 30 years later, Mill noticed the constancy of phenomena like the rates of murders, legitimate and illegitimate births, suicides and accidents. Even the number of letters discovered without an address at the London and Paris post offices was more or less constant from one year to the next. On the other hand, if we were to explain why some particular individual or set of individuals contributed to these rates, we would need a more individual kind of explanation: "...men's actions are the joint result of the general laws and circumstances of human nature, and their own particular characters; those characters again being the consequence of the natural and artificial circumstances that constituted their education, among which circumstances must be reckoned their own conscious efforts" (Mill, 1974:932).

Mill's adoption of Physics as his model for most of the social sciences, with the exception of economics which he left to the primacy of the Geometrical model, needs to be coupled with a clear understanding of his reservations about its literal application. Indeed, it is arguable that what he intended might be better represented as a metaphor than a model and that a great deal of wasted effort has resulted from an attempt to follow it slavishly. We also need to be clear that his image of physics is derived from nineteenth century science and does not necessarily reflect where physics has arrived in our own time.

Part of the problem is that Mill did not understand the idea of a law in the same way as a nineteenth century physicist would have understood it. As Ryan (1970:68–72) pointed out, Mill drew a distinction between what he calls empirical laws and laws of nature. The latter may be best understood as the small set of fundamental laws of physics from which the workings of the known universe may be deduced from the Big Bang onwards. Most empirical sciences, natural or social, deal with empirical laws. Even here, Mill varied between describing these as **summaries** of observations, 'everyone we have contacted who ate this meal has vomited'; as **low-level laws**, 'everyone, contacted or not, who ate this meal has vomited'; and as **derivative laws**, 'meals prepared under these conditions and containing these bacteria will cause vomiting'. When Mill talked about laws in social sciences, he usually meant the first or the second of these senses, whereas contemporaries like Comte, tended to see social scientific laws as more literally akin to those of physics.

In his emphasis on the deductive character of social science, Mill also tended to forget his own critique of deduction. Where did the principles or laws that he described come from? Ryan (1970:3–20) pointed out that Mill was sometimes thought to have been a member of the deductivist camp. In a modern formulation, Popper claimed Mill as an ally for his statement of the basic hypothetico-deductive account: "To give a **causal explanation** of an event means to deduce a statement which describes it, using as premisses of the deduction one or more **universal laws** together with certain singular statements, the **initial conditions**. (Quoted in Ryan, 1970:4.)

While Mill may have held a view of this kind in the 1820s, however, he had departed from it quite radically by the time *A System of Logic* was published. Popper thought that there was no 'problem of induction' because there was no such thing as 'inductive inference': Mill, on the other hand, thought that all deduction rested upon an inductive foundation.

Like Hume, Mill believed in the fundamental unity of scientific method and devoted quite a lot of effort to not very successful attempts to show that

mathematical and geometrical truths also came within this frame of reference (Ryan, 1974:66–70). As Ryan noted, the developments in both of these fields since Mill's work have made his arguments largely irrelevant, except in the sense of undermining the case that he was trying to oppose, namely that the deductive nature of truths in these fields made them models to which others should aspire.

However, Mill's objective was important in directing him to the basis on which deductive reasoning was claimed to be superior. He began by considering the classic form of deductive argument, namely the syllogism:

1. all men are mortal
2. Socrates is a man
3. therefore Socrates is mortal.

Given the major premise (1), and the minor premise (2), the conclusion (3) is necessarily true. But Mill asked, What makes (1) true? It is actually a summary of a set of particular statements: 'Hume is mortal, Mill is mortal...Socrates is mortal'. So the truth of the conclusion is actually a condition of the premise. We cannot be inferring the truth of something we know to be true already. Mill's conclusion is that general statements of the kind 'all men are mortal...' are actually summaries of particular statements and that the real inference is from these particular statements to the general ones. A general statement is best treated as a warrant or a justification for making an inference in a new case. Statements of this kind have been formed by induction from experience.

General statements in the form of laws of nature occupy a central place in explaining the natural and the social order. When a specific statement about causation is made, it implicitly invokes a general statement of some law. Ryan (1974:75) used the example of an injection of penicillin being said to cure someone's earache which appeals to a covering law of the kind that 'injections of penicillin cure earache'. But, following Hume, Mill also stressed that covering laws of this kind are not given in nature. The necessary connection that they express is founded on our experience. How, then, do we decide what sense to make of our observations such as to think that we may be justified in claiming that we have induced the operation of a general law or regularity?

Mill's great contribution was his answer to this question in specifying the canonical logical forms of inductive reasoning. A general statement could be shown to be true by one of four methods, which could be combined in five ways.

- *The method of agreement*

  If two or more instances of the phenomenon under investigation have only one circumstance in common, the circumstance in which alone all the instances agree, is the cause (or effect) of the given phenomenon. (Mill, 1973:390.)

  Suppose that we have two family planning clinics, which both fit 80% of users with IUDs. The clinics are held in different places, at different times of day, with different resources and draw users from different catchments. However, they both have the same doctor in attendance. The method of agreement would lead us to conclude that the cause of the similarity was the presence of that doctor.
- *The method of difference*

  If an instance in which the phenomenon under investigation occurs, and an instance in which it does not occur, have every circumstance in common save one, that one occurring only in the former; the circumstance in which alone the two instances differ, is the effect or the cause, or an indispensable part of the cause, of the phenomenon. (Mill, 1973:391.)

  Suppose that we have two family planning clinics, one of which fits all users with IUDs and one of which fits no users with IUDs. The clinics are held in the same place, at the same time of day, with the same resources and drawing users randomly from the same population. The first clinic is staffed by a doctor, the second by a nurse. The method of difference would lead us to conclude that the cause of the difference was the difference in the professional providing the service.

  These methods may be combined in the **Joint method of agreement and difference.**

  If two or more instances in which the phenomenon occurs have only one circumstance in common, while two or more instances in which it does not occur have nothing in common save the absence of that circumstance; the circumstance in which alone the two instances differ, is the effect or the cause, or an indispensable part of the cause, of the phenomenon. (Mill, 1973:396.)
- *The method of residues*

  Subduct from any phenomenon such part as is known by previous inductions to be the effect of certain antecedents, and the residue of the phenomenon is the effect of the remaining antecedents. (Mill, 1973:398.)

  Suppose that we take the same two family planning clinics. The nurse in the second clinic

establishes a job-share with a colleague who was trained at a different college. The nurse clinic now fits IUDs to 20% of its users. We have concluded from the induction above that the lower level of provision is due to the different professional backgrounds of the providers. We can now conclude that the actual level of provision by nurses is a function of both their general professional background and their specific prior training.

- *The method of concomitant variations*

  Whatever phenomenon varies in any manner whenever another phenomenon varies in some particular manner, is either a cause or an effect of that phenomenon, or is connected with it through some fact of causation. (Mill, 1973:401.)

  Take two more family planning clinics. One fits 80% of users with IUDs and the other fits 20% of users. In the former clinic, the doctor has four out of five sessions; in the latter she has one out of five. The method of concomitant variation would lead us to conclude that the variation in the rate of IUD fitting is caused by the variation in the number of sessions provided by the doctor.

As a number of commentators have pointed out, these formulations are not totally watertight, especially the method of residues, which is not easily distinguishable from the joint method of agreement and difference. Nevertheless, they do introduce a good deal more procedural clarity into this area than had previously existed and are still highly influential; much statistical inference takes similar forms. Multivariate analysis may be seen as a tool for the method of residues, for example. There are important practical difficulties: how do we know that we have considered all the possible similarities and differences, for instance? Mill certainly recognised this as a problem in extending these methods to social science (Mill, 1974:879–86; Fletcher, 1971b:216–21).

In his discussion of the relevance of the experimental method of chemistry, for example, Mill went through the canons of induction in relation to the example of whether free trade or protection better promotes national prosperity and showed that none of them can work with the necessary rigour. As both Ryan (1970:89; 1974:139) and Fletcher (Fletcher, 1971b:221) pointed out, however, Mill's argument is not entirely convincing, once we move away from the study of whole societies. If we ask whether a national health service (such as the NHS) is more effective than an insurance-based healthcare system at reducing rates of mortality and morbidity in a whole country's population, then Mill's criticisms have some force. If we ask whether nurse-led clinics offer better family planning services than doctor-led clinics, then we may have some chance of controlling other sources of variation, etc. sufficiently to approximate to the conditions required by Mill's approach, provided that we can agree on what 'better services' means. In general, the narrower and more specific the research question, the greater the likelihood of being able to carry out rigorous induction. However, we should also acknowledge that even a narrow question is vulnerable to Mill's critique, as we shall see later in discussing the experimental method in psychology, and that not all of the questions that are worth asking are narrow ones. We may very reasonably want to ask whether a national health service is more effective than an insurance-based system, even if this is a difficult question to answer.

The social sciences are founded on inductions from observation just as much as those of the natural sciences. While a strictly experimental approach may not be feasible, Mill's sequence – observe → induce → formulate → deduce → hypothesise → test → observe – is consistent throughout. The problem of achieving exactness in the social sciences is the complexity of the phenomena, their tendency to interact and to feed back. It is for these reasons that the social sciences cannot expect to achieve the degree of predictive power available in the natural sciences: the regularities simply do not exist to make this possible. Once we move away from a certain macro-level, society is a succession of unique occurrences. As far as the task of government or management is concerned, however, summaries and low-level laws are still better than reacting to every situation *de novo*. The manager who has some understanding of the principles involved in generating some particular state of affairs, and the ways in which they might interact with each other to create a situation that has not occurred in this precise form before and is unlikely to occur in this precise form again, is better placed to judge what lines of action are likely to be more or less effective and what information will be needed to assess and modify these if necessary. The successful manager, then, is unlikely to be a cook following a recipe book but a chef using a general understanding of the nature of ingredients and the available technologies of preparation, cooking and service to create and innovate.[38]

[38] See section 5.3.

At a macro-level, we can chart statistical regularities. If we want to know how they are produced, we need to look at actual cases. This is well brought out in the work of Harriet Martineau (1802–76), whose book, *How to Observe Morals and Manners* (1838), has a fair claim to being the first attempt to write in a self-conscious way about qualitative research. Martineau moved in the same circles as Mill and shared much of his philosophical stance, though he personally disliked her (Webb, 1960:11). According to McDonald (1994:168), she also stressed, at least in her early work, the unity of scientific method and its basis in observation and induction rather than deduction from *a priori* theories.

> [Bacon's] principle is so simple that it seems extraordinary that it should not sooner have been adopted ... This method is to bring together an accumulation of facts previous to the formation of a theory; and having carefully observed their bearing upon a particular point, to deduce [a slip] from them a principle which may be applied to the explanation of new facts. (Quoted in Webb, 1960:74 original editor's comment.)

She was later the first translator of Comte and her position seems to have shifted towards his more deductive approach, whereas Mill, after some initial enthusiasm, became progressively more sceptical.

How to Observe Morals and Manners begins with three chapters reviewing the general prerequisites for the traveller who expects to draw sensible conclusions from their observations. Martineau began by emphasising the need for direction in inquiry, caution in generalisation and objectivity in interpretation. In particular, she rejected the imposition of arbitrary standards of judgement or moral universals. Reflecting the direction which psychological thought had taken since Hume, she took it for granted that human characters were formed within specific societies as a result of the sensations to which people were exposed and that this process needed to be understood in its own terms (compare Webb, 1960:76–80). It is by this means that we may arrive at more fundamental notions: for example, human beings may everywhere seek to pursue happiness and to avoid pain but what counts as happiness and pain for a particular individual can only be understood in terms of a particular society and culture. The structure of a society and its values or culture are interdependent: the material circumstances of different members shape their values and vice versa. Information is best gathered by direct interaction with local informants and by the avoidance of excessive contact with one's own fellow-nationals. The prime sources of data are institutions and records. By institutions, Martineau was talking about the recurrent and routine activities of a society, which may be directly observed by the traveller. Indeed, she cautioned against putting too much weight on the evidence of informants except as an indication of the 'manners of discourse' adopted in the society. She then went on to discuss the kinds of institution that might be observed.[39]

One of these institutions is health. Martineau paid due deference to the importance of consulting civil registers and locally available statistics, and suggested that, if all else fails, much can be learned by visiting graveyards and noting ages of death. However, she insisted that health is an index of the morals and manners of the population rather than an item of interest in its own right: "Good and bad health are both cause and effect of good and bad morals" (Martineau, 1838:163). She rightly pointed to the limitations of partial local observations in forming universal plans for national preventive actions. The proportion of marriages to births and deaths, for example, provides a national indicator of the purity of morals and of the need for education to improve these. Variations between districts may help to identify particular areas that should be targeted. In the end, though, these statistical diagnoses must be coupled with a knowledge of the "prevailing occupations of the district to which they relate" (Martineau, 1838:165) to establish what means of education are required and the appropriate methods by which they may be delivered. Occupation is used here in its broader nineteenth century meaning of what people do in their everyday lives than in its modern sense of an element in an officially devised taxonomy of paid work.[40] Both **need** and **intervention** are identified inductively, she asserted; in fact, her approach looks rather more like the kind of *ad hoc* deductive use of statistical evidence criticised so fiercely by writers like Cicourel (1964) in the 1960s or Mill's inverse deduction. Martineau clearly had a prior theory about marriage rates as an index of morality which she used retrospectively to provide for a relationship or a necessary connection between the two.

Martineau (1838:231–8) concluded with a chapter on the means of carrying out an inquiry, which is still striking in its modernity; indeed it underlines

[39] See sections 4.2–4.6; [40] *Occupation* – what occupies one; a means of passing one's time, temporary or regular employment, business, calling, pursuit (*Concise Oxford Dictionary*).

the point with which we opened this chapter, that the basis of qualitative method is a very old one. The traveller needs to be prepared, to have organised a set of queries in advance, issues on which attention may be focused. This list, however, is provisional and must be constantly reviewed and revised in the light of emerging evidence about what is actually important in the circumstances under study.[41] It is not a list to thrust under the nose of informants or to dictate questioning but a means of ensuring a degree of systematicity in observing, asking and listening. In the field, it is important to record as factually, as regularly and as close in time to the observation as possible. Generalisation is best delayed until the completion of the journey, though it may be useful to note as speculation any theoretical ideas which are suggested at the time. There may be value in keeping two kinds of notebook – one for spontaneous observation in the field and one which is transcribed from this, expanded upon and containing the emerging theoretical ideas. This is almost precisely the same advice as Glaser and Strauss (1967) give and which would describe the practice of modern qualitative researchers in sociology or anthropology.

With the generation represented by Martineau and Mill we end what is, in effect, the pre-history of the social sciences in the English-speaking world. By the 1860s, economics has emerged in a recognisable form. It has shaken off Smith's comprehensive vision and settled to its classical axiomatic style of working, the Geometrical Method in Mill's terms. Comte has coined the term 'sociology' for some sort of scientific inquiry into society and this notion has begun to gain currency in the English-speaking world through Martineau's translation and Mill's advocacy. Herbert Spencer was about to start publishing his *Principles of Sociology* (Spencer, 1876), which would be the first major treatment of this field of study in English. Psychology has begun to move on from the introspective and theoretical debates prompted by Locke and Hume. History, which includes much that we would now call political science, has begun a methodological debate between those searching for laws of development and those who emphasise the idiosyncrasies of each age and the search for empathic understanding. Legal philosophy and jurisprudence were breaking away from the search for a fundamental natural law, with the same mystical properties as the elusive matter which Berkeley had repudiated, and to discuss law as a human creation designed to serve human purposes and capable of being evaluated in terms of its fitness in achieving them. Of the modern social sciences, only anthropology had yet to be given a definable identity that would clearly separate it from geography, on the one hand, and the tales of travellers, traders and missionaries on the other.

2.4 Towards a professional social science

In the British context, the major impetus for the development of qualitative methods in social research came from social anthropology as it developed after the end of World War I. This section will focus on that achievement and its impact on sociology towards the end of the 1930s. Towards the end, we shall look at the contemporary American picture and the emergence of the so-called 'Chicago School' of sociology. Although this did not have a major influence on British sociologists until the 1960s, this work, or a semi-mythical version of it, is the foundation of a good part of the qualitative research that is currently being done by sociologists in the UK.

Before looking at this, however, it may be helpful to sketch the broader methodological context. By the end of the nineteenth century, three kinds of social investigation had begun to emerge. The first, already referred to in relation to Mill, was the reanalysis of official statistics. This played an important part in the public health movement and in hospital reform. The second was the psychological experiment and the third was the social survey.

Friedman (1967) discussed the emergence of the psychological experiment in Germany in the 1860s out of astronomers' concerns for exactitude in measurement of stellar transits. Different observers of the same phenomenon recorded different times. On the one hand, this was a practical problem for improvements in astronomy, which the astronomers tried to address by minimising the human element in recording. On the other, it posed a problem of explaining the basis of human variation. Assuming that different observers were equally conscientious, why did they get different results? The phenomenon of reaction time had been identified and became the basis of a series of pioneering experiments to identify the effects of different attitudes and environmental

[41] See section 3.2.2.5.

conditions. However, at this stage a significant part was still played by the subject's introspection and self-reports. Friedman noted the implicit problems of experimenter effect on the reporting, an argument related to the issues about the status of interview data in other social sciences that will be discussed later.[42] It was only in the 1920s under the influence of Watson that the introspective elements disappeared and that was mainly because of their subjective elements rather than the possibility of experimenter effects. Watson was "quick to identify the employment of objective methods with the attainment of objectivity" (Friedman, 1967:15). Experimentation was reduced to a controlled stimulus and an observed response, preferably recorded by some automatic means. British psychology appears to have preserved a more social character for rather longer, at least to judge from Bartlett's work up to World War II. By the end of the War, however, experimental methods had established a dominance in psychology that would remain essentially unchallenged for the next 30–40 years.

The social survey is often thought of as the quintessentially British style of social research. Certainly it was pioneered by people like Booth, Mayhew and Rowntree, but similar kinds of inquiry were undertaken in most industrialised countries from the 1870s onwards. As Platt (1996) has stressed, however, it is important not to confuse the pre-World War II meaning of a social survey with the post-War questionnaire survey. The early studies predated the development of sample statistics in the study of agriculture and used various kinds of fieldwork assistant to collect data from everyone who could be identified as living in a particular neighbourhood.[43] These assistants might be middle-class volunteers, social workers, local officials or other people with some particular kind of interest in the problems being studied rather than the hired hands of the modern market research organisation. Their work often had a clear reformist objective of some kind. The data collection instruments did not have the degree of standardisation that we would now expect and the surveys regularly included significant qualitative elements, usually in the form of life-histories, which were used in the final reports to dramatise the quantitative findings. Although the best-known surveys are the urban ones, there were also regional surveys, linking the social life of a rural area to its topography, climate and ecology. These latter came to be the foundation of the discipline of geography and to play an important part in the evolution of anthropology. This broad sense of the social survey certainly continued until the 1940s and there were still aspects of it in the early community studies, discussed later in this chapter. Caradog Jones (1949:193) for example, presented a textbook of social survey work, which noted the impact of automation and standardised questionnaires but still saw a key skill of interviewing as 'the art of so directing the conversation that the questions are answered without any prompting on the part of the interviewer'.

2.4.1 The invention of social anthropology

In many ways, social anthropology can be seen as the first institutionalised research-led social science to be created in Britain. It went through a remarkable transformation in the thirty or so years that separated the publication of *The Golden Bough* (Frazer, 1890) and *Argonauts of the Western Pacific* (Malinowski, 1922). Frazer's book stands as the high-water mark of a tradition that goes back to the Ancient Greek historians, the collector and reteller of travellers' tales about the exotic ways of other societies. It was cast in an evolutionary framework, where these stories were used to examine the rise of modern civilisation and where the contemporary condition of non-western societies was treated as a proxy for our own prehistory. The aboriginal peoples of Australasia, Africa and America were treated, in effect, as living fossils whose culture could be treated as a guide to the early stages of our own historical evolution. The forms of kingship reported from these aboriginal societies were, for example, compared to those of the Ancient Egyptians or of the Hebrews. The essence of this work was its decontextualisation of the reported stories: all forms of relationship between king and clans, for example, would be treated as classifiable without reference to their meaning within the specific societies from which they arose. Frazer's work is typical of his time; Spencer (1874) used a similar approach to exemplify the laws of social development proposed by his sociology, as did Maine (1861) in his account of the creation

[42] See section 4.3; [43] Caradog Jones (1949) described the introduction of sampling methods into social surveys in Britain. Although Bowley used this approach in a study of Reading, then a small market town, in 1912, it clearly did not become fully accepted until the mid 1920s. The 1928 New London Survey, which was partly designed as a successor to Booth's work in 1886–87, used both Booth's method of total population interviewing and Bowley's sampling with the intention of validating the latter approach and testing the potential generalisability of Booth's original conclusions.

of law. However, as this generation sought to formulate more general statements about the nature of societies and their development, the fragility of such a method was becoming apparent. Travellers – traders, missionaries, planters or whoever – did not necessarily ask the right questions and their reports lacked the context that was required to take social science onwards to the study of societies in an holistic fashion (Urry, 1993:17–40).

The next generation took up the challenge of collecting their own data by going to these traditional societies in person and questioning native informants, usually through an interpreter. The first of these is usually regarded as the Cambridge University Expedition to the Torres Straits (between Australia and New Guinea) in 1898–99, organised by Haddon and including Seligman, who did further work in Melanesia, India and the Sudan at various times over the next 30 years, and Rivers, who went on to other work in the Pacific before World War I (Stocking, 1995:98–115; Urry, 1993:73–7). Haddon and Rivers's student, Radcliffe Brown, worked in the Andaman Islands in 1906–8 and in Australia during 1910–13. In the USA, Boas began working in the Pacific Northwest of the USA and Canada in 1886 and made 13 visits to the area until 1931 (Rohner, 1966:151–212), though, reflecting the greater linguistic element in early American anthropology, he sought to become capable of working in the native languages. The experience of the Torres Straits work persuaded the participants that a different kind of method was required. Haddon, who had begun his career as a zoologist, began to talk about something called **field work** (Stocking, 1995:114). It was, however, Rivers, a physician, who did most to define this, especially through his work with the committee to revise *Notes and Queries,* a manual that had first been produced by the British Association for the Advancement of Science in 1874 to guide data collection by amateur anthropologists. The book was a list of questions that could be asked by missionaries, district officers, magistrates or travellers that would contribute to the cumulative study of humankind, which was still not clearly distinguished between sociology and anthropology. *Notes and Queries* was revised in 1892 and 1899. In 1907, a committee, including Haddon, Seligman, Rivers and a fourth member of the expedition, Myers, was appointed to revise it again. The new edition was published in 1911(Stocking, 1995: 17–40; Urry, 1993:120–3). Rivers's programme is, perhaps, most economically summarised in a report that he wrote in 1913 on anthropological work outside America for the Carnegie Institute:

> A typical piece of intensive work is one in which the worker lives for a year or more among a community of perhaps four or five hundred people and studies every detail of their life and culture; in which he comes to know every member of the community personally; in which he is not content with generalised information, but studies every feature of life and custom in concrete detail and by means of the vernacular language. It is only by such work that one can fully realise the immense extent of the knowledge which is now awaiting the inquirer, even in places where the culture has already suffered much change. It is only by such work that it is possible to discover the incomplete and even misleading character of much of the vast mass of survey work which forms the existing basis of anthropology. (Quoted in Urry, 1993:28.)

Survey work is used here in the way discussed earlier. It is not a survey in the modern sense of a questionnaire administered to a sample population but, rather, a reasonably systematic effort to collect data about some particular group or locality from a mixture of official sources, professional informants – in this case, traders, teachers or missionaries – or by the structured questioning of native informants.

Rivers is notable as a man whose thinking changed and developed throughout his life and his approach to field work appears to have been no exception. What is not clear from the literature that it has been possible to consult is the extent to which he was able personally to follow through this programme, having begun as an orthodox questioner of informants. The most persuasive case is made by Barker (1995) in her reworking of his field notes from his final expedition to Melanesia in 1914 for her novel *The Ghost Road.* She seems to have used a source not examined by others and which Rivers did not live long enough after World War I to write up – he died in 1921 after making important contributions to psychiatry during the war.[44] As a result, the credit for pioneering what we would now call participant observation has largely gone to the charismatic

[44] The relations between anthropology, psychiatry and psychology before World War II could not be explored further in the time available for this review. There is, for example, an interesting discussion of the relationship between Radcliffe Brown and Bartlett, who was one of the major figures of British psychology between the Wars, in the latter's book *Remembering,* which reports on some of his classic experiments with memory. Much could also be said about the relationship between anthropology and psychoanalysis, though this was less close in Britain than in the USA.

Malinowski, whose account of his own work in Melanesia is published in *Argonauts of the Western Pacific* (1922).

As we have already seen, the climate in anthropology was already moving in this direction. Stocking (1995:119) identified at least nine other students of Haddon and Rivers, at Cambridge, or Seligman and Westermarck, who was Professor of Sociology, at the London School of Economics, who had all undertaken field work before World War I. Malinowski, however, designed *Argonauts of the Western Pacific* as a manifesto and it begins with an introduction of startling clarity about the methodological programme that would dominate British and, to some extent, American anthropology until the 1970s.

Malinowski (1922) distinguished three main principles: any researcher must espouse scientific values, must live among the people being studied and must apply a number of special techniques for collecting, ordering and presenting evidence.

His first injunction was that "the results of scientific research in any branch of learning ought to be presented in a manner absolutely candid and above board" (Malinowski, 1922:2). He went on to note that "it would be easy to quote works of high repute, and with a scientific hall-mark on them, in which wholesale generalisations are laid down before us, and we are not informed at all by what actual experiences the writers have reached their conclusion" (Malinowski, 1922:4). He emphasised the need to distinguish clearly between the results of direct observation, native statements and native interpretations, on the one hand, and, on the other, the inferences drawn by the researcher.[45]

Fieldworkers should not seek out the company of their own kind but should remain in sustained contact with their subjects. This allows them to become part of the ordinary life of the people they are studying so that their presence ceases to disrupt routine activities. Fieldworkers gather data through direct observation and spontaneous discussion which they then use to induce statements about the regularities which constitute their subjects' culture. These statements are, in turn, tested through systematic questioning. Such abstracted statements must, however, remain linked to what Malinowski (1922:18) called **the imponderabilia of actual life**. In this he is contrasting the formal bonds of the social group with the flux of everyday interaction which constitutes the strongest and most subjectively real sense of the community in the eyes of the subjects. Members of the society experience it as a series of routine and everyday encounters and not in the abstract inductions of the anthropologist. Here again, direct observation is central to the work:

> There is no doubt, from all points of sociological ... analysis and in any question of theory, the manner and type of behaviour observed in the performance of an act is of the highest importance. Indeed behaviour is a fact, a relevant fact, and one that can be recorded. And foolish indeed and short-sighted would be the man of science who would pass by a whole class of phenomena, ready to be garnered, and leave them to waste, even though he did not see at the moment to what theoretical use they might be put. (Malinowski, 1922:20.)

Through this attention to the everyday routine, the part played by particular events like ceremonies or rituals can be given its due salience. This is a specific challenge to what might be called the **Frazer approach**, where ceremonies and rituals would be decontextualised and compared with each other as isolated units of analysis: "the fundamental obligation [is to give] a complete survey of the phenomena, and not [to pick] out the sensational, the singular, still less the funny and quaint" (Malinowski, 1922:11). Corresponding to accounts of typical events must go accounts of the typical ways of thinking and feeling about those events among members of the society. Here an understanding of the native language and the way in which it is used to organise and classify events is vital. Indeed, Malinowski advocated the collection of a corpus of native-language accounts which could be inspected by others as a semi-autonomous verification of the observer's analysis. The final goal, for Malinowski, was "to grasp the native's point of view, his relation to life, to realise **his** vision of **his** world" (Malinowski, 1922:25 original emphasis).

Although there is an important history of theoretical debate within what had come to be called social anthropology – to distinguish it from anthropometry or physical anthropology, though not yet from sociology – the basis of Malinowski's methodology went unchallenged for half a century. Whether students of Malinowski or Radcliffe Brown, the other major figure of the period, the core of anthropological study was the period living among a small community of non-Western people. As Kuper has commented:

[45] See section 5.3.

> After Malinowski, the anthropologists based their methods upon participant observation which required intimate and free contact with the peoples they studied. They therefore had to break down the barriers of the colour bar, which existed in most colonies, and they had to challenge the assumptions of all colonial regimes. Their individual examples of how sophisticated Europeans could happily adopt many tribal habits and live on a basis of friendship with illiterate and poor people constituted a constant irritation to settlers and many colonial officers. (Kuper, 1975:149.)

The tendency of modern writers to separate fields and disciplinary histories in ways that would not necessarily have made sense to contemporaries has rather obscured the overlap between anthropology, sociology and, to some extent, geography right through to the 1960s. Although the use of qualitative methods was most clearly articulated in anthropology, their application was not limited to non-Western populations. Haddon worked with the British Association for the Advancement of Science to try to stimulate an anthropological survey of the British Isles in the 1890s (Stocking, 1995:104–5; Urry, 1993:83–101). Malinowski endorsed the Mass Observation programme of the late 1930s (Mass Observation, 1938).

2.4.2 Mass Observation: anthropology comes home

Mass Observation was an informal organisation created by a group of artists, film makers and intellectuals in the 1930s who were looking for a way to revitalise socialist and democratic politics in the context of the rise of fascism and the stultifying effects of a conservative government with an unshakeable majority and a lack of any vision about how to deal with the deep economic recession. Mass Observation was intended to by-pass the establishment channels of orthodox politics and media by giving ordinary people a chance to express their views and opinions and to have these brought to national attention. The organisers used two main methods of achieving this. One was to recruit a nationwide network of volunteers who kept diaries and in which they recorded their own daily life and reaction to public events and noted what people they knew, encountered or simply overheard, were thinking and talking about. Sometimes the volunteers were asked to stimulate discussions on particular topics or to listen particularly for reactions to certain events. The other approach was to send teams of people to particular places for fairly short periods of time to make an intensive study of talk and behaviour for a specific purpose. Blackpool in the summer was selected for one such investigation into working-class leisure, for example. The reports and diaries were collated in London and turned into a series of published volumes which attracted quite extensive coverage in national press and magazine reviews. Although Mass Observation is often treated as a literary or cultural phenomenon, or simply lost in the survey agency that it eventually transmuted into, its founders clearly had a vision of a qualitative sociology, an 'anthropology of civilised societies', for which they were ultimately unable to elicit institutional support.

Malinowski's contribution to Mass Observation's report on its first year's work may well be the first direct confrontation between advocates of qualitative and quantitative methods.[46] Despite its amateurish beginnings, Malinowski was eager to claim the work as a "natural out-growth of the best tendencies in modern social science" (Mass Observation, 1938:85). In particular, he offered the group a robust defence against those who have criticised their work as unscientific. The founders of Mass Observation had responded by trying to quantify their observations; Malinowski thought they should be defended on their own terms:

> What is science? In substance, it is nothing but reliable fact, presented in a sound theoretical argument: science is observation based on sound judgements. The further paraphernalia of scientific work: figures and graphs, statistics and charts, may be useful; indeed when they are legitimately used, are indispensable; when not they are the pseudo-scientific rig-out often covering integral fallacies. The only test of science is whether its generalisations are based on ascertained fact and carried out with unimpeachable logic. (Mass Observation, 1938:89–90.)

Science is a systematic process of observation, inference and integration into a general theory. Its objectivity is a matter of attitude rather than specific techniques. The subjective behaviour of human beings is the objective data of the social scientist. Fieldwork is not a matter of empathy, though it may be one of sympathy. The anthropologist's respect for the people he is studying is an important element of his methodology. However, they remain objects of his appreciation.

> The distinction here is fundamental. By learning a strange language we acquire the code to the understanding of expressive acts which, since they are symbolical, remain completely meaningless without

[46] See chapter 3.

> their practical mastery. But is it necessary in order to understand your best friend, to have made love to his wife? ... In the same way the Ethnographer who studies, for instance, cannibalism, is not under any obligation to eat missionary chop. Why? Simply because the only difference between Mutton and Missionary is perhaps in an infinitesimal difference of taste, which is quite irrelevant to any sound scientific treatment of the subject. (Mass Observation, 1938:99–100.)

Social science faces a fundamental confusion between the personal reality of acting in the human drama and the outside registration of acts as a clue to the inner process. Qualitative methods are a means of accomplishing that registration. Malinowski had particularly clear responses to people who question the repeatability of observations. On the one hand, major field sites are regularly revisited and inconsistencies examined; on the other, sciences such as astronomy, vulcanology and meteorology often make use of unrepeatable observations to assemble bodies of data whose uniqueness facilitates the test of existing generalisations. Language is central to this as a "tool, a document and a cultural reality at the same time" (Mass Observation, 1938:106). The reports of the Mass Observers to the central office are akin to those of informants in anthropology. They constituted texts for analysis by those with a more scientific training and should be treated as data in this respect rather than simply being aggregated in the way that they had been in the first year's work. Mass Observation would progress by moving on from its empiricist foundations rather than by defensively reformulating them in quantitative expressions.

Malinowski's comments were clearly taken on board by Tom Harrisson, one of the prime movers behind Mass Observation. Harrisson's most coherent methodological statement was published in a paper, *The Future of Sociology* in 1947. Harrisson began by denouncing "the philosophical approach to sociology, in which great laws of human behaviour are produced without observation" and the reaction to this in the form of "an absorption in quantitative methods which, with maximum impersonality, satisfy purely 'mathematical' criteria" (Harrisson 1947:10). He suggested that this gap can be filled by an approach which aims to ascertain laws of social behaviour as a result of studying human relations and institutions in the natural condition. (Like Haddon, Harrisson was trained originally as a biologist who then extended his field studies to human societies.) This would parallel the work of anthropologists on so called 'primitive' peoples; sociology was the anthropology of civilised societies. Harrisson highlighted the neglect of four issues; verbal behaviour; the observation of social behaviour, the penetrative analysis of social institutions; and the intensive study of social problems. This results particularly from the obsession with statistical method as the only alternative to philosophy. His critiques anticipated many later and better-known accounts. Questionnaire responses have a problematic relationship to actual behaviour; they are better treated as evidence of attitudes or what people think they ought to say rather than as evidence of what they do. Bias is inherent in all observational and coding schemes and is not eliminated by the use of a mathematical language to conceal it. Objectivity is no more guaranteed by one person's design of a survey instrument than it is by that same person's observations. In each case, it is the explicitness of the methodological description that allows others to judge and, if appropriate, to seek to replicate and check the results.[47]

Harrisson concluded by listing and evaluating seven types of method: questionnaire interviews; indirect and informal interviews; overheards; case and life histories; documentation; observation; and penetration. These are mostly self-explanatory, except for overheards, where he was talking about systematic listening to unprompted conversations, and the last two which refer to different locations for observation – public and private settings. American social science, he suggested, was making more rapid progress than British because of its greater willingness to embrace the variety of methods available for the systematic study of social life.

The use of American models is particularly important, and may well reflect Malinowski's own close connections with the University of Chicago. Harrisson cited as his exemplars the *Middletown* studies by Lynd and Lynd (1929; 1937), the Chicago studies inspired by Park and Burgess (see section 2.4.3) , the work of Dollard on race relations in the American South, Warner's *Yankee City* studies (e.g. 1963) and the Western Electric (Roethlisberger and Dickson, 1939). Studies by Mayo's team from Harvard. The inspirations of these studies drew variously on sociology, psychology and anthropology and all of them tended to mix a variety of qualitative and quantitative methods. However, as Harrisson rightly noted, the qualitative dimension was central to them in a way that could not be identified in British work of the same period. He did not exempt Mass Observation itself from these strictures; while he had been away on war service in Borneo, where

[47] See section 5.3.

he had pursued sociological and anthropological studies in the early 1930s, Mass Observation had become increasingly influenced by government contracts and begun its transition to a conventional market research organisation. One of its last studies in the traditional style was actually a look at the impact of the introduction of the NHS on general medical practice. This was probably the first self-conscious attempt at qualitative medical sociology in the UK.

Meet Yourself at the Doctor's (Mass Observation, 1949) is a short (70 page) account. The first chapter looks at what would now be called illness behaviour and the second and third chapters report popular views of the NHS before and after its introduction. Both of these topics are examined on the basis of panel members' reports. Chapter 4 reports the views of two general practitioners (GPs) in partnership, one of whom favoured the NHS and the other of whom opposed it. This is followed by an account of a typical working day for another GP, which includes some observational data. Chapter 6 is based on 'overheards' from the waiting room of a practice serving a working-class area in north-west London, and chapter 7 is based on panel members' accounts of their interaction with doctors in both general practice and hospital settings. There is then a description of observation in a child health clinic, focusing particularly on the work of the health visitors and, finally, a chapter looking at the reported use of patent medicines and marginal healers, and at lay beliefs about illness prevention.

At about the time this paper was published, Harrisson returned to Borneo where he was government ethnologist in Sarawak until 1966. As a result, his work had little direct impact on British sociology. However, his essay is important in demonstrating that the major American models of qualitative research were available more or less as they were published in the 1930s and 1940s, even if they had little impact on academic sociology. Harrisson also looks forward to the development of anthropological research on communities in Britain, which we shall examine later in this chapter. The history of British social science would almost certainly have looked very different if Harrisson's arguments had received more attention at a time when institutional development was beginning, in the wake of the Clapham Committee's 1946 Report on the infrastructure for social and economic research. As it was, these American writings had a patchy influence until the 1960s when the expansion of the higher education system and the demand for social research to inform policy making under the Wilson governments brought a rapid enlargement and increasing professionalisation of the social science community. With these increases in scale and new policy and intellectual challenges, there was more space for heterodoxy in the face of the continuing dominance of the philosophical and empiricist approaches that Harrisson had targeted.

2.4.3 Qualitative research in America 1920–1960

As Platt (1994; 1996) has stressed, the history of sociological research in the USA has been substantially mythologised. According to this myth, qualitative methods, particularly participant observation, were invented at the University of Chicago, were its almost exclusive form of research and were not done in any significant way anywhere else before World War II. In fact, all of these statements are untrue. Qualitative research was not uniquely associated with the University of Chicago: important studies were done by people from other institutions throughout the period.[48] The data used were often second-hand and taken from personal documents or social casework records rather than from direct observation or interviewing. Quantitative work was always undertaken and valued. Nevertheless, the force of the myth has some substance behind it. Chicago sociology and, to a slightly lesser extent, anthropology were the dominant organisational and intellectual forces in those fields from the early years of the century until the late 1930s. For much of this period, a qualitative version of economics in the form of the institutional school associated with John Commons at Chicago held a similar sway, though this was giving way to the neo-classical approach of Frank Knight and his colleagues at Harvard by the late 1920s.[49]

[48] What seems to be the first major treatment of health and health care in sociology, for example, forms part of the *Middletown* studies carried out by the Lynds, who were most closely associated with Columbia University in New York, which is often portrayed as the quantitative antithesis of Chicago. These two studies of Muncie, Indiana were carried out in 1925 and 1935 and provide an important document of the institutionalisation of medicine and the emergence of the hospital as its centre of practice in this period. The Lynds drew heavily on anthropological thinking, especially in the first book, and acknowledge Rivers's programme as a significant influence; [49] An account of the qualitative tradition in economics would be a major undertaking from original sources. However, there is clearly an important aspect that could be added to this review tracing the history of Austro-German economics in parallel to that of neo-Kantian sociology from Böhm-Bauwerk through von Mises and Hayek.

The Chicago School has been the subject of a substantial body of scholarly work, which can only be summarised here (Faris, 1967; Carey, 1975; Dibble, 1975; Rock, 1979; Bulmer, 1984; Shore, 1987; Fine 1995). Much of its character derived from the new understanding of the mission of universities introduced more widely with the foundation of the University of Chicago. In contrast to the older East Coast universities, Chicago rejected the traditional English model of the university as a means of transmitting a high culture to a social elite in favour of the German model of the university as a centre for research, innovation and the rational education of professional, managerial and technical workers. Small, who was recruited to found the Sociology Department in 1892 shared this vision. While most of his own work had the theoretical emphasis that was typical of his generation, he stressed the importance of such thinking being developed from and tested by research. Sociology, as he wrote in his 1894 textbook, was essentially an empirical discipline based on "the method of observation and induction" (Small and Vincent, 1894:15). Direct observation was the key both to objective knowledge and to an appropriate moral concern: "There is little likelihood that men who personally observe actual social conditions, according to the method we propose, instead of speculating about them in their study, will want to fold their hands and let social evil work out its own salvation" (Small and Vincent, 1894:374).

However, Small's own contribution was mainly as an organisational leader and the key influences on the development of research appear to have been Thomas, who worked in the department from 1895 until 1918, and Park who was there from 1914 until 1933.

Thomas established the links with anthropology, which developed within the sociology department until 1929 and remained close until the 1950s. His massive study of migration, *The Polish Peasant in Europe and America,* with a Polish philosopher, Znaniecki, has long been regarded as one of the classic empirical studies of sociology (Thomas and Znaniecki, 1927). The data on which this was based were mainly documentary – self-reported life histories, family correspondence between people in Poland and their relatives in the USA, newspaper files, immigration records and casework reports. As Blumer, a later recruit to the department, commented at a conference to review this research in 1937, the actual use of the materials did not measure up to the standards of the study's Methodological Note in terms of the demonstration of a specific relationship between data and theoretical conclusions. The authors also provided very little information to allow others to make judgements about the representativeness, adequacy, reliability and validity of the data. Nevertheless, the study and its methodological discussions had a strong influence on the practice of sociological research in the USA throughout the inter-war period. As Platt (1996) has recently noted, the classic Chicago studies of the early 1920s drew rather more on these kinds of documents than subsequent legend has suggested. The students took much of their data from life-histories, which might or might not have been collected directly, and from the casework reports of the various social work agencies that had close associations with the department.[50]

However, Platt is almost certainly too dismissive of the amount of direct observation that was undertaken, particularly under the inspiration of Park. He had been a successful investigative journalist in the 1890s but came to be dissatisfied with the *ad hoc* nature of such inquiries and to seek a more systematic framework for understanding societies. The central role in this was played by direct experience and observation. Two of his students recalled his injunctions:

> Park made a great point of the difference between knowledge about something and acquaintance with the phenomenon. This was one of the great thrusts in Chicago, because people had to get out and if they wanted to study opium addicts they went to the opium dens and even smoked a little opium maybe. They went out and lived with the gangs and the ... hobos and so on. (Carey, 1975:155.)

> One thing more is needful: first-hand observation. Go and sit in the lounges of the luxury hotels and on the doorsteps of the flop-houses, sit on the Gold Coast settees and on the slum shakedowns; sit in the Orchestra Hall and in the Star and Garter Burlesk. In short, gentlemen, go get the seat of your pants dirty in real research. (Bulmer, 1984:97.)

One might speculate about the conversations that Park and Malinowski could well have had on the question of whether anthropologists were required to eat Missionary Chop, if offered!

[50] See section 4.4 for a discussion of the use of documents in contemporary research.

In methodological terms, the Chicago tradition was largely an oral one. Bulmer (1983) noted that the classic monographs of the 1920s contained very little discussion of their methods or methodology. There was a brief account of observation in Vivien Palmer's (1928) *Field Studies in Sociology*. According to Bulmer, this book mostly concerns itself with general principles of social research and does not go into details about specific practices. However, Bulmer's essay introduced a previously unpublished manuscript by Paul G Cressey (1983), written about 1927, in which he described his work on the study that became *The Taxi-Dance Hall*. Taxi-dance halls were places frequented by men who were marginal to working-class society in Chicago, either because of transience, ethnicity, race or disability. The men bought tickets with which they paid women working for the hall to dance with them. The halls had a reputation for low-level vice – prostitution, drug-dealing, mugging, etc. – and were a target of social reformers. Cressey first became involved as a caseworker for a charity concerned with juvenile welfare. However, as he showed, this role only gave him certain kinds of access. It was useful in interviewing the parents or relatives of the taxi-dancers themselves, especially to elicit factual information in short interviews. However, this obstructed his access to the halls themselves by contrast with a colleague from a marginal ethnic group, who found the hall-owners happy to talk to him about most topics, except the women. Cressey found that his only way in was to learn how to pass as a member of the group, to learn the special language and presentation of self required of male patrons. In doing so, it was essential to suspend moral inhibitions and pre-judgements. He tells the story of a young woman fellow-student who obtained work as a dancer to help with the project and failed completely to establish rapport because of the way in which she immediately began to make moral judgements on her informants. Cressey discussed the problems of recording data, of eliciting factual information without an obvious reason for doing so, of the hierarchy of topics that can be talked about between strangers and about the methods of verifying accounts:

> Against the objection that in the opportunity for 'expansiveness' and fantasy which the role of anonymity makes it possible for the subject to affect is itself 'untruthful' and 'inaccurate', many replies can be made. Not the least important is that in the long run, a story if told in bits and essentially truthful throughout tends to 'hang together'. Each new aspect of the person's story as it comes out dovetails with the rest. Where the person is lying, this dovetailing does not go on. ... Finally, it may be said that sociology is more interested in attitudes, wishes and conceptions of the world, and the life-story which is not essentially correct in every detail may be more valuable by reason of the fact that these inner aspirations and fantasies are revealed. (Cressey, 1983:119.)

Cressey's last remark introduces an important theoretical contribution of the Chicago work: its interest in how people perceived the world and acted on the basis of their perceptions rather than in establishing some objective truth, which most of the scholars involved eventually came to accept was unknowable. Thomas and Znaniecki formulated this in *The Polish Peasant in Europe and America* in the idea of the **definition of the situation**. They saw this as the basis of a 'social technology' that would allow the more purposeful control of social situations. The technologist would need to combine three kinds of information: the objective material conditions; the actors' prior knowledge, values and experience; and the actors' working definitions of what kind of event they were engaged in and the ways in which conditions, knowledge, values and experience were relevant. As Thomas later summed up the position, in one of the most famous aphorisms of the discipline of sociology, "If men define situations as real, they are real in their consequences".[51] Znaniecki (1963:242–8) discussed a number of examples: the farmer whose actions are based on his expectations of the weather, of the condition of his machinery, of the likely state of the market at the point of harvest and so on. This knowledge is necessarily imperfect – there may be a sudden hailstorm, the combine harvester may break down, an international trade agreement may lead to a reduction in prices or whatever. What is interesting for the sociologist, however, is the way in which people do their best

[51] The quotation actually comes from a book called *The Child in America,* which is rather like a systematic review of research on behavioural problems in children and evaluations of interventions designed using various models of human behaviour. Although the book was jointly authored by Thomas and his wife, Dorothy Swaine Thomas, she attributed authorship of this quote to him in a letter to Robert Merton. It is worth noting that the sentence occurs at the end of a paragraph discussing the delusion of a murderer who killed people at random in the street because he saw their lips moving and thought that they were saying insulting things about him. In the next paragraph, Thomas retreats from the radically subjectivist implications by stressing that human behaviour also needs to take account of the objective material features of situations as well as of the beliefs about these which motivate actors. This context is often ignored when the quotation is used.

on the basis of imperfect knowledge and what the results are.[52] The farmer's actions reflect his beliefs and its expectations – the typical course of weather in his area over the summer months, the general condition of his equipment, the usual trends in the market for his produce. These expectations may not be realised in the future but they are the necessary and inevitable foundation of his actions here and now. It is the difference between using medical audit as a means of censuring people for not achieving the auditors' standards and as a starting point for questions about why this particular pattern of processes and outcomes has come to occur as the result of people doing what they believe to be the right things. In a sense, it takes us on from the positions of people like Descartes and Berkeley in recognising that there may be a loose fit between our understandings of the world and the world itself but, as Weber saw, our understandings are objective data on the foundations of action which have practical importance and can still be studied in their own right.

Social organisation is the outcome of these local actions based on local resources and knowledge. In Park's words:

> [Societies] are formed for action and in action. They grow up in the efforts of individuals to act collectively. The structures which societies exhibit are on the whole the incidental effects of collective action. (Gee, 1929:8.)

Park, however, also recognised the difficulty of the Thomas and Znaniecki project in a way that their original approach did not acknowledge. They had tended to assume that the documents they had examined in *The Polish Peasant in Europe and America* could be taken more or less at face value as evidence of the attitudes and values of the authors. Park pointed out that these are actually inaccessible to us. People's mental states have to be inferred from their actions: "We get the meaning of social institutions as we get the sense of words, by observing the ways in which they are used; by investigating the occasions and incidents of their origins and growth, and by taking account of whatever is unusual or unique in their history" (Gee, 1929:37–8).

These problems of inference were elaborated in the philosophical work of Mead, who was working in Chicago at the same time as Park. Mead developed a social psychological theory of **social behaviourism**, which he contrasted with the stimulus–response models of psychology that were being developed contemporaneously in the Psychology Department at Chicago. The radical behaviourism associated with Watson and, later, Skinner reduced all action to operant conditioning that associated stimuli directly with responses. Intermediate notions like 'mind' were dismissed as metaphysical and unscientific; all that could be observed were programmed reactions to inputs detected by the five senses. Mead argued that, while this account might fit animal behaviour, such as the rats and pigeons used in the classic experiments, it did not fit human behaviour. Sensations did not have inherent meanings. One of the things that we learned as children was how to experience our bodies. How did we recognise and label sense-data? Lindesmith and Strauss (1949) later used this idea to examine heroin addiction. From the point of view of social action, they observed, the important moment was the addict's discovery of the relationship between the physical sensations of withdrawal and the relief offered by the drug. Although this could be described as a biological process, it was the recognition of the connection and the realisation that the sensations could be altered by taking the drug that formed the basis of the addict's actions in seeking to obtain and use the substance. Unpleasant as the symptoms of withdrawal might be, their definition as 'addiction' and their prompting of a search for the drug rested on the addict's interpretation of the experience and the ways of finding relief. There was nothing inherent or intrinsic about the conversion of these physical experiences to social action. Becker (1967b) later developed these ideas in relation to other psychoactive drugs such as marijuana, LSD and corticosteroids.

As Mead pointed out, many of the sense–data that we receive are symbolic, requiring decoding and interpretation to establish what they meant. An often-used example is that of an old-fashioned school classroom. A pupil raises his hand. The teacher's observation of this is sense data. However, she now has to decide what these data represent or symbolise. Is the pupil volunteering to answer a question, requesting permission to leave the room, relieving a writing cramp or what? She must then consider whether to acknowledge the behaviour

[52] The problems of action in the real world of imperfect information are an important element of the Austro-German tradition in economics, as opposed to the theoretical world of perfect information assumed by the neo-classical approach.

by collecting further information, perhaps by asking the pupil what he wants. Having done that, she still has to consider the status of the information. If the pupil wants to go to the toilet, does the pupil need to do so or is it an attempt to disrupt the lesson or to manufacture an opportunity to smoke an illicit cigarette or even some combination of these? A process of this kind cannot be captured as a set of conditioned reflexes.

Again, however, one can note the oral nature of the tradition. Mead died in 1931 at the age of 68 having published relatively little: most of his work in print has been reconstructed from manuscripts or students' notes. Mead's work had relatively little impact on Chicago sociology in his lifetime but became important through Blumer, who wrote his PhD on Method in Social Psychology in 1928 and was on the staff of the Department of Sociology from 1925 until 1952. It seems to have been Blumer who coined the term **symbolic interactionism** for the kind of sociology that was directly inspired by Mead and which had a major impact after World War II through its influence on people like Shibutani and Strauss.

However, as a number of accounts stress, it is important not to overstate the social psychological aspects of Chicago sociology. Shore (1987), for example, noted the influence of plant ecology, which was also a major intellectual strength in the biological sciences at Chicago, on Park's thinking about the social organisation of cities. The botanist Frederick Clements had developed an analysis of plant formations which argued that they were arranged in a succession of zones, which displayed an evolutionary sequence: higher life-forms displaced lower ones by invading existing areas and pushing the lower forms to the edge of their original habitat. Plant distribution showed the succession of colonisation in a sequence of zones of settlement. Clements argued that the arrangement was only partly determined by environmental conditions such as soil moisture or acidity: plants also modified the habitat to suit themselves. Eventually, one species would achieve a stable dominance because of its adaptiveness to the environment and its ability to resist incursions. This species would only be dislodged by some external change such as a climatic shift or the import of plants from an alien ecosystem. In effect, Clements saw the plant community as a kind of organism in its own right, which passed through a developmental process similar to that of any individual organism. It had its own definite laws and structures which were, nonetheless, spontaneously realised.

Social organisation might very often be an unintended consequence of collective action but the ecological metaphor provided a way to study it without reference to specific individuals. This links back to evolutionary traditions in nineteenth century social thought and, in particular, to Spencer, the English social theorist. Social organisation could also be studied at a level above that of individuals as a matter of competition for resources and of interdependencies which result from this: "By this means the numbers of living organisms are regulated, their distribution controlled and the balance of nature maintained ... the existing species – the survivors in the struggle – find their niches in the physical environment and the division of labour between the species develops" (Park quoted in Shore, 1987:102).

This perspective became known as **urban ecology**. It was an important point of connection with the work in human geography which had developed towards the end of the 19th century and which, through regional surveys, also played some part in the shaping of social anthropology: Boas, for instance, began as a geographer. As Shore (1987) noted, Ratzel, writing in the 1880s and 1890s, had defined geography as a study of the relationship between human beings and their physical environment and Vidal de la Blache had first drawn attention to the possible model of plant ecology in his *Principles of Human Geography,* first published in 1922. Park argued that the difference between human geography and urban sociology was that the former was interested in the relationships between people and their physical environment while the latter studied the relationships between people as they were mediated by their physical environment. Chicago urban sociology saw the city as a set of ecological niches, each with its own distinct social characteristics, that were successively colonised by different migrants. A slum would tend to remain a slum but to be occupied by the most recently arrived ethnic group, who would gradually move up and out as they reorganised themselves under the new conditions of the city.

Although the literature of urban sociology would have important implications for health service delivery in many community and primary care settings, it has rarely been used for this purpose. Its importance for the present review is in the way that the model was further extended, particularly in the work of Hughes and his students to study the organisation of work and occupations. Hughes completed his doctorate in 1928 on the way in which the spatial organisation of the city of Chicago was sustained by the workings of the Real Estate

Board. This provided a crucial bridge between the study of the city and the study of work. Although Hughes continued his interest in urban issues while working at McGill University from 1927 until 1938, when he returned to Chicago, his subsequent work shifted towards the study of occupations. Hughes conceived of the work that needed to be done in a society as analogous to the landform of a city, with hills, rivers, marshes, etc. that were crucial physical constraints on development but which came to be incorporated by the social organisation of the urban community. The land was divided socially in particular ways by the workings of the local culture and the local economy. Some districts became slums, some became suburbs. The transitional zones between them were the sites of competition and struggle. In the same way, some work was dignified and some was dirty, both physically and morally. Some tasks were the subject of competition for ownership, others were neglected. Occasionally, technical change would create new areas, just as the reclamation of the marshes along the Chicago river or the in-filling of the shores of Lake Michigan created new development opportunities in Chicago. Where the city had a socio-cultural structure of residential and commercial neighbourhoods, the world of work had a socio-cultural structure of occupations, the division of labour. Just as ethnic groups competed for control of the material and symbolic capital of the city, so too groups of workers organised to compete for control of the material and symbolic capital of work.

As Hughes (1971:418–21) himself noted, the end of World War II saw an influx of bright graduate students from a wide range of previous occupational backgrounds who were interested in understanding the impact of the social, cultural and technical changes brought about by the war. Quite a number of these had backgrounds in health care, which was a part of the world of work that was experiencing particularly dramatic changes. As some of the critics of Chicago urban ecology had pointed out, the model was very dependent upon the specific dynamic characteristics of a rapidly-growing city experiencing massive immigration. In certain respects, the hospital after World War II had a similar dynamism. The accelerated technical developments of wartime brought a rush of change as they were introduced to civilian populations at the beginning of a long economic boom that provided constantly growing resources for investment. It became a natural laboratory for the Second Chicago School's distinctive sociology of work and occupations (Fine, 1995). In the 15 years from 1920 to 1935 only one out of 71 PhDs (1.4%) from the department had focussed on a health-related topic and that was a quantitative study of mental disorder. Between 1946 and 1961, 11 out of 217 PhDs (5%) had a health-related theme. Although the numbers were still small, the influence was substantial, especially when added to some of the research projects like the study of the socialisation of medical students in Kansas that became *Boys in White* (Becker *et al,* 1961), which brought together Hughes, Becker and Strauss with the 'outsider' Geer, who had a PhD in Education from Johns Hopkins University. Others, notably Eliot Freidson, moved into the health field after PhD studies on different topics – Freidson's was on mass communications.

The Chicago studies had a distinctive perspective on medical work and medical organisations which they tended to approach 'from below'. *Boys in White,* for example, is often contrasted with the contemporaneous study of medical education by a team from Columbia University, *The Student Physician* (Merton *et al,* 1957), which is seen as having a much more establishment perspective. The Kansas students are strategic actors, trying to work out how to survive in a complex organisation that makes unbearable demands on them. The Columbia students are more clearly embryonic doctors, compliant absorbers of the values and culture of their profession (Atkinson, 1983). The same spirit was carried into studies of patients – Goffman's *Asylums* (1961), Roth's *Timetables* (1963), which looked at TB patients, and Davis's *Passage Through Crisis* (1963), which studied families where a child had polio. Patients struggle to make sense of the organisations and the professional actors who confront them. In the mental hospital, the inmates create an underworld of their own where they show a remarkable degree of initiative and creativity in managing the strange and impoverished environment that surrounds them. In the TB ward, patients construct their own benchmarks of progress from one day to another. Are they getting better or worse? When will they be allowed out? Allowed home? They observe the hygiene rituals of the staff. Why do some professionals wear masks and others not when performing exactly the same task? The polio victims and their families have the same problem of judging recovery. How much progress have they made? How much will they make? The genuine clinical uncertainties of prognosis are compounded by the functional uncertainty engendered by the professionals, the theory that compliance with treatment can be better encouraged by keeping families hopeful by stressing the unknowns beyond the point when

the doctors are quite certain of the likely outcome. These studies retain an important contemporary relevance because of the way in which they deal with fundamental questions about the social organisation of health care which tend to recur in new forms wherever new technologies arise. The distinction between functional and clinical uncertainty, for example, recurs in recent studies of genetic counselling (Bosk, 1992) and neo-natal intensive care (Guillemin and Holmstrom, 1986). However, their impact was delayed in the UK by a variety of intellectual and material barriers.

2.4.4 Qualitative research in Britain since World War II

Harrisson's comments on the state of British sociology at the end of World War II, while displaying his renowned acerbity, were not far off the mark. Although one of the early post-war issues of the *British Journal of Sociology* carried Goffman's first published paper, *Symbols of Class Status* (1951), the discipline remained dominated by, on the one hand, debates on social theory, particularly between those who had been affected by the formalist thinking of Tascott Parsons from the USA and those who were influenced by various versions of Marxist thought from the European mainland, and by quantitative survey work of one kind or another. The latter often stood at the peculiarly British boundary between sociology and social policy.[53] It focussed on high level issues about the nature of social stratification and social mobility in modern societies, at a time when the occupational and class structures were undergoing rapid change. The growth of the NHS and other welfare bureaucracies, for example, created rafts of new white-collar and professional jobs that could only be filled by a degree of social mobility that had not been seen for many generations. Who were these people? How did they get to be where they were? What did it mean for their values and for traditional relationships between social classes? Another important and related area was in the study of educational achievement where much important work was carried out on the influence of social class and on the limitations of 11-plus selection for different tracks through secondary school. There was also a burgeoning of work in criminology, attempting to identify predictors of delinquent behaviour. Some of these issues overlapped with developments in social medicine and the emergence of social epidemiology in the investigation of the relationship between various health indicators and social class. This became an important route into health-related research for a number of sociologists in the 1950s, many of whom were to become prominent figures in the field of medical sociology by the end of the 1960s. With the time and resources available for this review, however, it has not been possible to identify any qualitative research on health-related issues in the UK from Mass Observation's (1949) *Meet Yourself at the Doctor's* until the 1970s.

Other social sciences went through a similar experience. Both psychology and economics also became dominated by formal theories and by quantitative methods. Only in anthropology and, to a lesser extent, geography did a qualitative tradition survive in UK studies, through the vehicle of what has become known as community studies. The classic survey of this body of work was prepared by Frankenberg (1966), who had himself produced one of the major contributions in his study of a Welsh village in 1953 (Frankenberg, 1957). Frankenberg's review examined 13 studies of communities in the British Isles from the work of the American anthropologists Arensberg and Kimball in the South West of Ireland during the 1930s through the work of Rees (a geographer), Williams and Frankenberg (anthropologists), Birch (a political scientist), Stacey, Young and Wilmott (sociologists) and concluding with a varied group of studies of new housing estates, mainly by sociologists and planners. The list of authors alone gives some sense of the diversity of the work, though Frankenberg made an heroic attempt to synthesise it through his notion of a **morphological continuum** reflecting the varying density of social relations under different environmental conditions. Although this line of work continued into the 1960s, it gradually became less fashionable as the difficulties of characterising entire communities became more apparent and social science itself became more specialised. Stacey's (1960) original study of Banbury in 1948–52, for example, was conducted by the author, with a fairly small amount of time from two research assistants, and produced what appeared to be a comprehensive account of life in this small town. The re-study from 1966–69 involved a team of four, all with a professional education in sociology or anthropology, and was

[53] With the exception of one or two Old Commonwealth countries, this distinction is not found anywhere else. Applied research on social problems or issues of policy is either conducted in 'mainstream' departments of sociology or, in the form of public policy, political science or in interdisciplinary institutes which bring together people with a range of more specialised social science training.

much more focused on issues about social stratification and economic life. It also made much greater use of questionnaire methodology in an attempt to get a more representative sample of the community as a foundation for its generalisations (Stacey *et al,* 1975; see also Bell, 1977). In more recent years, the tendency has been to ask more specific questions and to use community as a site for investigating some other topic rather than seeking to characterise it in a holistic fashion. A recent example might be the work by Greenhouse and co-workers (1994) on dispute resolution in three American towns where law is seen as a resource for the construction of what it means to feel part of a community. The interest is more in the **idea** of a community than in seeing physical proximity itself as important (Crow and Allan, 1994). Indeed, we now find discussions of **virtual** communities, people linked by means of technology, and of **transnational** communities, people from particular social, cultural or ethnic origins who are globally dispersed but retain a discernible social coherence.

Although these community studies had a certain influence on thinking in public health (they featured prominently, for example, on reading lists for health visitors who were required to do a neighbourhood study as part of their training) and both Frankenberg (1992) and Stacey (e.g. Stacey *et al,* 1970)[54] later made distinguished contributions to medical anthropology and sociology, there is very little that is explicitly health-oriented in this body of work. The major exception is Cornwell's (1984) study of health and illness in Bethnal Green in the late 1970s. This was based in a geography department but was a self-conscious attempt to revisit aspects of Young and Wilmott's (1962) classic work. However, its primary objective is not the study of community *per se* but on the use of the concept of community to understand the social distribution of health and illness in a traditional working-class neighbourhood. What did people themselves understand the nature of their bodies and their disorders to be and how was this knowledge shared, transmitted and reinforced between them? What were the implications of its intersection with the knowledge and delivery systems of officially-recognised allopathic medicine? Were there ways in which the disjunctures contributed to the statistically observable regularities in the distribution of mortality and morbidity in this geographically-defined area?

Put thus, Cornwell's work provides a bridge to a much wider body of work in medical sociology and anthropology on lay understandings of health and disease and the ways in which these may help to account for actions that appear irrational by the standards of official medicine. We shall return to this shortly. For the present, the important point to retain is that the main contribution of community studies was in keeping alive the idea of qualitative methodology during a period of considerable pressure to emulate the successes that were apparently being achieved by quantitative approaches and in offering opportunities for people to acquire some of the practical skills involved. The reinvigoration of qualitative work in the late 1960s came from other sources. Ironically, these were in the shadow of precisely those areas where quantitative methods appeared to be achieving their greatest successes.

One of these was criminology, where the Home Office had sustained both a well-funded in-house research unit and a substantial academic community since the end of World War II. This had resulted in a steady flow of survey work, psychometric studies and secondary analysis of official statistics in an attempt to steer policing and criminal justice policies and to identify causes of crime. In the political atmosphere of the late 1960s, the limitations of this agenda were sharply exposed. Crime, it was argued, was not a neutral and objective phenomenon but the result of the ways in which laws defined certain acts. These acts were not randomly distributed in a society but reflected the situational responses of certain individuals and groups to their social environment in ways which those who controlled the law-making process found threatening. Similarly, criminal statistics were not objective maps of the incidence of crime. They demonstrated the selective attention of law-enforcement agents and skewed processes in the systems of prosecution, adjudication and disposition. The young sociologists who articulated these criticisms drew rather indiscriminately on a rediscovered Chicago tradition of qualitative work and on various strands of Marxist class analysis – the exact mixture varied between authors (Taylor *et al,* 1973). The main forum for these ideas was the National Deviancy Conference (NDC) which met irregularly from 1968 until 1972 and whose general approach is exemplified by the *Images of Deviance* collection edited by Stanley Cohen (1971). This movement reformulated the core intellectual

[54] Gordon Horobin, who played an important role in training the 'Aberdeen School' of qualitative researchers, also worked on a community study of Hull fishermen in the early 1960s.

question as the study of deviance rather than of crime. What did a society define as deviant behaviour? What kinds of deviance were there in a society? What were the implications of classifying one way rather than another? Who did the classifying? How was the classification imposed?

Although the NDC collapsed in disarray, largely as a result of political quarrels, this movement had a considerable impact on the development of a qualitative medical sociology in the UK. This influence was partly intellectual and partly personal. Intellectually, the sociology of deviance picked up the traditional criminological interest in mental illness and turned this around with a combination of resources from the anti-psychiatry movement represented by people like Laing (1965) and Szasz (1972) and from more mainstream symbolic interactionist research like that of Scheff (1963). If mental illness was the product of societal definitions of normal and abnormal behaviour, could the same arguments not be made about physical illnesses? Consequent upon this was the idea that the whole definitional structures of medicine might be studied in the same way as the definitional structures of criminal justice. The notion of physical illness as a form of social deviance had been well-established in the USA for some 20 years (Parsons, 1951) and had inspired a stream of writing which was pulled together at a critical moment by Freidson (1970) in his award-winning work, *Profession of Medicine*. The virulence of the conflicts between old and new criminologists also had the critical impact of denying the latter access to the funds and patronage which were largely controlled by the former. A number of sociologists of deviance drifted into studies of medicine, partly because jobs were available and partly because there was a clearer separation between the intellectual and the political agendas. Strong was probably the most influential of this group.

The other area where there was somewhat less visible conflict but which made an important contribution was the sociological and anthropological study of schools. The extensive body of work on 11-plus selection had identified the relationship between social origins and educational achievement without being able to explain how this came about. One response was to ask questions about the organisation, curriculum and pedagogy of schools (Hargreaves, 1967; Lacey, 1970; Young, 1971). Intellectually, this had much in common with the sociology of deviance, with the established institutions being seen as, in effect, creators, sustainers and reproducers of the problems of which they complained. Working-class underachievement was the product of an educational system which reflected the choices that a society, or at least those in power, had made about the knowledge, values and attitudes to encourage and to sanction or repress. Methodologically, this area of work has been very influential, as will be seen by the number of citations to people like Hammersley and Burgess later in this review. However, there have been more direct influences on the study of medical institutions and medical work, most notably through the contribution of Atkinson (1981), who was trained as an anthropologist and came into medical sociology through a project on undergraduate medical education in Edinburgh, which drew considerable inspiration from *Boys in White* and from Becker's later writings on the interactionist studies of schooling (Becker, 1970). Indeed, the work of Atkinson and his colleagues has probably been as influential on the sociology of education as on the sociology of medicine.

At the beginning of the 1970s, British medical sociology could be divided into two main schools. The **London** school was dominated by graduates of Bedford College, trained in the social administration traditions of Jefferys (e.g. Jefferys, 1965) or the quantitative social psychology of Brown (e.g. Brown and Harris, 1978). The **Provincial** school was a looser network numerically dominated by the Medical Research Council Medical Sociology Unit in Aberdeen but also involving significant groups in Edinburgh and South Wales. London was identified with survey analysis, 'social factors in ...' contributions to epidemiology and with a traditional Fabian political agenda. The Provinces were identified with qualitative work, with a more theoretically informed approach, with a less deferential stance towards medicine and with a less self-conscious political position. In retrospect, these divisions were probably never as sharp as they seemed at the time: Aberdeen also produced a great deal of well-regarded social epidemiology, though London was slow to accept qualitative studies. The personal political beliefs of the Provincials were not very different from those of the Londoners, though they were more deliberately separated from their academic writings. However, the differences were sufficient to sustain separate labour markets and the relative immobility of the academic profession over the last 25 years has meant that these tendencies retain a surprising degree of visibility.

As we approach the present and lose some of the perspective of history, it becomes more difficult to describe the patterns. However, if we consider the course of development in qualitative medical

sociology since 1970, two major points of evolutionary division can be discerned. The first took place in the 1970s between those who became influenced by a new wave of thinking from the USA known as ethnomethodology. The other, which occurred at the end of the 1980s, responded to the European ideas of post-modernism.

In order to understand the reception of ethnomethodology by British medical sociologists, it is important to appreciate the small scale and isolated nature of the community around 1970–72. There were probably no more than 30 recognisable qualitative researchers, very few of whom had ny direct contact with the sociologists from the USA whose work in medicine and related areas was being taken as a model. In effect, the British taught themselves from US books and journals, with some practical support from people who had done qualitative work within community studies. The result was to homogenise very different intellectual traditions, which had different histories and sources and which were, to a degree, in competition or conflict within the USA.

Ethnomethodology is an approach to the study of social life founded by Garfinkel, who received his PhD from Harvard in 1950 but has spent most of his career at University of California, Los Angeles. Although some of the post-war graduates from Chicago, or from Brandeis, where Hughes moved after his retirement from Chicago, went to the West Coast, Garfinkel's approach was separate to an extent that did not become fully clear to British readers for some years after 1970. The term ethnomethodology was coined to describe a kind of investigation that would focus on "the ways in which ordinary people ('ethno') methodically constructed their social world". It emerged from Garfinkel's debate with his early Harvard mentor, Parsons. The latter was associated with a systems approach to sociology where individual social actions were seen to be derivative of some set of rules that specified how people should behave in order to sustain the order of the society. Parsons's work was often associated with positivist styles of empirical research, though this is not intrinsic to his position and some of his students produced qualitative studies, of which the most notable body of work in the study of health care is probably that of Fox (1974). Garfinkel challenged this model, arguing that people were not rule-followers, whose behaviour was the manifestation of some kind of social programming, but rule-users, constantly trying to work out how to match the specificities of particular situations to whatever general orientations might be helpful in deciding how to act. Some of the classic case studies which emerged from this focussed on the social construction of statistical data. It emerged that the coding of data was not an unproblematic application of the coding rules but a constant process of trying to decide what the coding rules ought to mean in a particular case (Cicourel and Kitsuse, 1963; Garfinkel, 1967; Cicourel, 1968; Cicourel, 1974; Buckholdt and Gubrium, 1979[55]; Gubrium and Buckholdt, 1982). As a result, the statistics reported did not represent a literal description of the incidence of the phenomenon but were mainly evidence of the way in which commonsense classification practice operated in a particular situation. One of the simplest examples can be found in Atkinson's study of decision making by coroners about how to count a death as a suicide, where one operating in a seaside resort described how deaths by drowning were likely to be considered suicides if the deceased had left their clothes folded on the beach and accidents if the clothes were in a heap. As Atkinson pointed out, road traffic deaths were never scrutinised as possible suicides despite the suggestions in the literature of a possible association between adverse life events and accidents. Decisions like these 'filled in' the official legal definition of a suicide but also generated the statistics on which suicide research rested (Douglas, 1967; Atkinson, 1978). More recently, Prior has looked at the practice of death certification and its implications for mortality statistics. Relatively unskilled coding clerks make commonsense decisions on which great conclusions rest (Prior, 1985; Bloor, 1994[56]).

This is, and remains, a crucial challenge to the validity of the statistical data on which much health service research is founded. It was a powerful and attractive tool in the hands of the provincial medical sociologists who were seeking to challenge the dominance of quantitative methodologies, both in the London version of their field and among the epidemiological and public health communities in whose shadows they worked. However, the critique also extended to most of the data generated by social scientists, whether by surveys, attitude measurement, interviews or observation (Cicourel, 1964). Social scientists also produced models of the way in which people organised their lives which reflected the common sense of their disciplines. Sociologists, for example, were preoccupied with

[55] See appendix 1 for details of this study; [56] See appendix 1 for details of this study.

mapping so-called deeper structures of class, gender, race or whatever on to their findings in order to tell a real story, which explained people's actions in ways that were unrelated to the production of those actions in the contexts in which they occurred. The main force of this attack was felt in sociology and, to a lesser extent and mainly in the USA, anthropology, though in the 1990s it has also been acknowledged by psychologists. In his study of the transsexual, Agnes, Garfinkel placed an emphasis on discovering when and how a feature like gender was demonstrably relevant to people in their actions that has served as a guiding principle ever since (Garfinkel, 1967; Schegloff, 1997).

These arguments had a profound effect on the way in which some medical sociologists did qualitative research. They abandoned the idea of *verstehen*, which had been central to both symbolic interactionism and to the post-Weberian understanding of social interaction. It was no longer thought possible to have access to the meaning of events for the actors by some sort of empathic or imaginative act. 'Taking the role of the other', putting oneself mentally in the position of a person being studied risked importing a bundle of sociological preconceptions without specific evidence to justify this. However, if sociologists could not access the minds of the people they were studying, neither could those people themselves access each other's. In the absence of telepathy, we as a species had to have evolved practical ways of demonstrating the mindedness of interaction and procedures for checking and repairing problems in this process. Those practices and procedures could become the data for a social science. The result was a shift in research focus to a much more specific analysis of language, body movement and social interaction. In the process, a division opened up between those who retained an emphasis on recovering meaning, who were accused of over-interpretation and the selection of evidence to fit prior theses, and those who shifted to the study of practices and were, rightly, thought to be rather indifferent to the moral or other implications. The differences in approach are explored in more detail in chapter 6.

A further wedge was driven into the qualitative community by the spread of post-modernist thinking towards the end of the 1980s. This movement from cultural and literary studies was felt most strongly in sociology and anthropology but has had a certain impact on geography and, latterly, in psychology. Superficially, post-modernism has a certain affinity with ethnomethodology and, indeed, some of the most radical early ethnomethodologists espoused positions which were recognisably its precursors. Post-modernism also emphasises the local production of the sense of a situation and the theoretical impossibility of imposing a definitive reading. There are as many versions of a literary or cultural text as there are readers, as many versions of a social situation as there are participants or observers. Some of these are privileged over others and one objective of a post-modern analysis is to explain why and to promote the acceptance of alternatives. This tends to introduce a strong political agenda, the silenced versions usually turn out to be those of historically oppressed groups. It is this agenda that brings observations together in any form of generalisation, which is otherwise normally eschewed. The findings of a post-modern study are simply one more version, which is not any way privileged over any other. Traditional criteria for the evaluation of work by reference to its reliability and validity, its sampling procedures, its distinction of fact and opinion and so on, are treated primarily as rhetorical sleights of hand by which those with the power to privilege some versions of the world conceal that power behind apparently objective and impersonal criteria.[57] Many more traditional qualitative researchers have been unhappy with the comprehensive demolition of realism implied by these arguments. (Silverman, 1989; Silverman, 1993; Silverman, 1997; Hammersley, 1992a; 1992b; 1992c).[58] Although some of the people influenced by ethnomethodology have had some sympathy with the post-modern case, it is probably correct to say that most have tended to see its radical social constructionism as a trivial position. Human beings are embodied entities who live in a material world, both of which constrain the possibility of versions. We can to a great extent modify our material world and to some extent modify our bodies but we cannot do either indefinitely. In theory we can behave as we choose interactionally; in practice to do so risks a variety of negative inferences and sanctions. What is interesting is not the theoretical indeterminacy of the world but its practical determinacy. The return to hyperbolic doubt seems unlikely to advance our understanding of anything.

Although this account of the last 30 years has concentrated on sociology, most other social sciences have also seen some kind of revival of qualitative methods. Medical anthropology has developed as a distinct sub-field and there has

[57] See section 5.2; [58] See section 3.2.1.1.

been some qualitative work in cultural and human geography. Within the broad field of social and public policy and administration studies, which spans academic activities that take place in departments of social policy, political science and management, there has also been a revival of this approach through what Glennerster and co-workers (1983) have called **administrative anthropology**. Psychologists have arrived more recently at the party as the dominance of the experimental paradigm that had ruled more or less continuously since World War II came into question (Harré and Secord, 1972). Economics is, perhaps, the only social science to have been left untouched: the revival of interest in the institutional school and the Austro-German work of the 1920s and 1930s has yet to produce a return to its methods, except to the extent that some of the case study work in management schools may be seen as a partial success.

Implications for commissioning and practice

- Qualitative methods are not a recent invention in the social sciences, though their application in the field of HTA and HSR may be novel. This means that they should not be treated as controversial merely on the grounds of being new or different from the traditions of HTA or HSR. However, it also means that both commissioners and researchers need to benchmark research in these fields against these well-established external standards if they are to avoid either reinventing wheels or supporting, disseminating and acting upon second-rate science.
- This survey of almost 400 years of methodological debate underlines the complex and difficult problems of ontology and epistemology which arise in all scientific inquiry. Although these often frustrate practical men and women who would like to see all research findings reduced to checklists, bullet points or score cards for action, they are fundamental to understanding the limits of confidence to our ability to have knowledge of and to act upon the world around us. A distinctive feature of qualitative research has been its willingness to accept and to confront these problems, even at the cost of underselling the robustness of its own claims to representativeness, reliability and validity.
- This section also brings out the ultimate indivisibility of the human sciences. Although these have split into a variety of specialties and subspecialties, in much the same way as medicine or surgery, an holistic approach to evaluation problems is not impossible. The barriers tend to reflect the particular paths that particular disciplines have followed through the methodological debates of the last hundred years or so and which have become institutionally dominant in different fields. Effective multidisciplinary working, of the kind often required in HTA, is, however, unlikely to be achieved unless team members are prepared to address these debates on a basis of mutual respect. This makes it difficult to create teams 'off the shelf', as members are unlikely to be willing to make these compromises in the absence of a longer-term professional incentive, rather than staying safely within the existing boundaries of their disciplines in order to maximise their chances of future employment and advancement. The depth of knowledge required to participate in these debates also tends to tell against people with a formally multi-disciplinary training, as this often leaves them too superficially equipped to become involved, though it may give them appropriate skills for commissioning work or utilising the results.

Chapter 3

The relationship between qualitative and quantitative methods

Qualitative research in HTA is largely concerned with the part people play in healthcare interventions. People are involved in such interventions as providers and recipients of existing treatments, services and new technologies. Measurement of healthcare need and approaches to prevention of disease relate specifically to people, whether at the individual, group or population level. It is difficult to think of health technology without thinking about people. In that sense HTA is a social science. As chapter 2 has shown, the development of the social sciences has been marked by a series of debates about the nature of the social world and the methodologies which are appropriate for its study. In recent years, this debate has centred upon the relative merits of qualitative and quantitative methods and the appropriate relationship between them. It is to this that we now turn.

In setting out to compare and contrast qualitative and quantitative methods, we are faced with the problem that there is a singular lack of consensus about the fundamental tenets of either tradition. Particularly among advocates of qualitative methods, there are profound differences of opinion about the nature of the scientific enterprise, the extent to which social research can or should aspire to be scientific, the ways in which the social world can be studied and the criteria which should be applied to the products of such study. We have not progressed far from the situation described by Bernstein in 1976:

> The initial impression one has in reading through the literature in and about the social disciplines during the past decade or so is that of sheer chaos. Everything appears to be 'up for grabs'. There is little or no consensus – except by members of the same school or sub-school – about what are the well-established results, the proper research procedures, the important results, the important problems, or even the most promising theoretical approaches to the study of society and politics. There are claims and counter claims. (Bernstein, 1976:xii cited in Lincoln, 1990:68.)

Sandelowski (1986) claimed that qualitative research is an imprecise and misleading term, since each qualitative method has its own rules, aims and methods. While it is possible to identify common elements it is difficult to generalise in a way that does justice to the various traditions and approaches: "The debate surrounding the methodological rigour of qualitative research is confounded by its diversity and lack of consensus about the rules to which it ought to conform and whether it is comparable to quantitative research" (29).

The contested nature of qualitative research is highly relevant to those seeking to commission or carry out research in the field of HTA. Different versions of qualitative research exist which may be more or less appropriate to the goals of HTA. Thus qualitative research cannot simply be adopted or dismissed *en bloc*. Given the lack of agreement among qualitative researchers themselves about what it is for, how it should be done, or how it should be evaluated, it is meaningless to argue for or against qualitative research in general. We must unpack the assumptions which underpin the practices of those advocating one or other version of qualitative research if we are to make judgements about the usefulness of such approaches to research in HTA. In particular, greater precision is required if we are to establish whether it is possible to combine qualitative and quantitative methods in such research.

There is a similar lack of consensus about the assumptions which underpin the use of quantitative research methods. It has been conventional to characterise such methods as **positivistic**, though the exact meaning of this term is disputed (Halfpenny, 1979; Lincoln and Guba, 1985; Bryman, 1988). However, it has also been argued that it is inaccurate to describe all research within the quantitative tradition as positivist and all within the qualitative tradition as anti-positivist (Reichardt and Cook, 1979; Bryman, 1988).[1] A significant difference between qualitative and quantitative methods is that, while the latter have established

[1] See section 2.3.2.

a working philosophical consensus, the former have not. This means that quantitative researchers can treat methodology as a technical matter. The best solution is one which most effectively and efficiently solves a given problem. The same is not true for qualitative research where proposed solutions to methodological problems are inextricably linked to philosophical assumptions and what counts as an appropriate solution from one position is fatally flawed from another. In this section of the report we will map out the different positions adopted by qualitative research methodologists.

The consideration of the relationship between qualitative and quantitative methods, in this section of the report, falls into two parts. First, we shall outline the various positions which have been adopted by researchers in relation to the appropriate relationship between qualitative and quantitative methods (section 3.1). We shall then go on to consider some of the grounds on which some qualitative researchers have argued that qualitative and quantitative research are fundamentally different enterprises, characterised by different philosophical assumptions, methodological principles and practices (section 3.2).

3.1 Qualitative and quantitative methods

The overall lack of consensus among qualitative researchers is perhaps best illustrated by the range of views expressed in the literature about the relationship between qualitative and quantitative research methods. The extent to which qualitative and quantitative methods may be seen capable of collaborative co-existence is hotly contested both by those advocates of quantitative methods who dismiss qualitative approaches as not proper science and enthusiasts for qualitative methods who reject quantitative approaches as doing violence to the social world. Duffy (1987) has described this as the separatist versus combinationist debate (131).

Leaving aside those who reject any role for qualitative research, the debate about its role in furthering knowledge can be seen as taking three broad turns. Before going on to clarify these three positions, it is important to acknowledge that they are primarily a heuristic device. In practice, elements of the three positions are frequently conflated and, as Bryman (1988) has argued, advocates of qualitative methods tend to vacillate between them.

3.1.1 Choosing between methods on instrumental grounds

There are those who argue that the choice between qualitative and quantitative methods should be made entirely on instrumental and pragmatic grounds. The question to be asked can be expressed as 'Will qualitative or quantitative methods provide the answers to this question most effectively and efficiently?' Those who advocate the choice between methodological approaches on such instrumental grounds tend to argue that quantitative methods are inadequate, **on their own**, to meet our research objectives, particularly when the research involves people, and are more likely to tolerate and even advocate combining of quantitative and qualitative methods in a joint enterprise and to see each approach as contributing something to our overall knowledge of a particular field. Qualitative researchers starting from this position may criticise the over-enthusiastic application of quantitative methods to research objectives which would be better served by a qualitative approach. They do not, however, condemn quantitative methods, in principle.

Qualitative and quantitative methods are seen as different and potentially complementary ways of gathering data, whose usefulness depends upon their appropriateness for a given research task. For example, McKinlay (1993) has described qualitative and quantitative methods as "mutually enriching partners in a common enterprise" (113). The appeal here is to the technical superiority of qualitative methods in at least some circumstances (Silverman, 1989). This position was summed up by Holman (1993:35): "True understanding in medicine cannot be achieved without adding qualitative to the research arsenal".

Hammersley and Atkinson (1995) argued that, before the 1940s, qualitative and quantitative methods were used side by side.[2] They suggested that it was only with the rise of logical positivism with its absolutist tenets, that qualitative research began to be conceptualised as an alternative paradigm.

Within this broad instrumental position, three slightly different perspectives can be discerned, and these are discussed below.

[2] See chapter 2.

3.1.1.1 Qualitative research as junior partner

In the first place, there is the view of qualitative research as a somewhat unfortunate necessity, which offers a solution when preferred quantitative methods are impractical. For example, Dean and co-workers (1969) commented:

> As scientists we want to be as rigorous as possible. Whenever a crucial experiment or survey will provide data of testing relevance for our theories, we will want to use them. But there are many areas of social science where this cannot be done. (274)

They saw a particular role for unstructured methods in the exploratory and pilot stages of research, where the research question is not yet clearly formulated or relationships which need to be examined have not yet become explicit. Such methods may be used to generate hypotheses, which can then be tested using quantitative methods. However, they would accept, there are some situations in which quantitative data is fundamentally inaccessible. For example, carrying out a sample survey of people involved in illegal or stigmatised activities might well pose major practical difficulties. In such cases, qualitative methods may have to be relied upon, not only for the generation of hypotheses, but also for their testing. Finally, they presented unstructured methods as particularly appropriate for an intensive study of an individual case. Such individual cases might include the intensive study of an innovative intervention, prior to its wider introduction.[3] While these authors identified a number of strengths of qualitative methods, they also noted what they saw as fundamental weaknesses – their inappropriateness for statistical treatment and the likelihood of bias arising from the researchers' close involvement in the setting or group they are studying. Qualitative methods are clearly seen as coming fairly low in the authors' hierarchy of methods. As they commented in a footnote: "The refinements of the experiment still serve as the model towards which we strive" (Dean *et al,* 1969:21 footnote).

More recently Imle and Atwood (1988) have identified a role for qualitative methods in supporting the development of sensitive quantitative instruments. They made clear that, for them, measurement is the priority and qualitative research is reduced to a handmaiden role.

3.1.1.2 'Horses for courses'

A second group of researchers specifically eschew any attempt at a hierarchy of methods. The following quotation from Trow (1969) sums up this perspective:

> It is with [the] assertion that a given method of collecting data – **any** method – has an inherent superiority over others by virtue of its special qualities and divorced from the nature of the problem studied, that I take sharp issue. The alternative view ... is that different kinds of information about man and society are gathered most fully and economically in different ways, and that the problem under investigation properly dictates the methods of investigation. (332 original emphasis))

Trow's position is a popular one. (See also McCall and Simmons, 1969; Vidich and Shapiro, 1969; Zelditch, 1969; Walker, 1985; Stange and Zyzanski, 1989; Waitzkin, 1990; Steckler *et al,* 1992; Holman, 1993; Munhall, 1993c; Baum, 1995; Secker *et al,* 1995; Silverman, 1997). It is essentially a 'horses for courses' argument. The analogy of the tool-kit is also frequently used. In similar vein, Baum (1995) has referred to a 'smorgasbord of methods', from which the researcher should choose, on the basis of which method is likely to produce the most comprehensive and valid answers. Trow argued that it is simply inappropriate for the surgeon to ask whether the scalpel is a better instrument than forceps. The scalpel is only better if the task in hand is cutting. Likewise, the researcher should not argue about whether qualitative methods are better than quantitative methods, without adding the rider, 'better for what?' Silverman (1997) suggested: "There are no principled grounds to be either qualitative or quantitative in approach. It all depends on what you are trying to do" (14).

Munhall (1993c) argued that qualitative and quantitative methods should be seen as complementary and mutually reinforcing. She envisaged a cyclical relationship between qualitative and quantitative research. Qualitative research was seen as a first level, descriptive activity, while quantitative research permits the testing and verification of models which are based upon the thick description of such qualitative work. However, she argued that once a probabilistic relationship is uncovered between two variables, using quantitative variables, further qualitative research is needed to understand the exceptions which do not fit the probabilistic conclusions. In this way the model can be further refined.

From this perspective, it is unhelpful to see any particular method as the gold standard against which others are to be measured. Vidich and

[3] See chapter 7 for a discussion of the role of qualitative research in programme evaluation.

Shapiro (1969) argued for an alternative approach to the evaluation of research findings, in which different methods are treated symmetrically and the limitations of every method of social research is acknowledged: "Unfortunately any alternative procedure [to participant observation] for the study of culture and social behaviour is also subject to serious questions of validity" (295).

All research findings are to be assessed in relation to the extent to which they offer "a correct account of [a group's] attitudes, values and practices" (295). Alongside this criterion of informational adequacy, Zelditch (1969) added a second, that of efficiency. Best methods are defined as those which are seen as efficiently yielding the most adequate information. Similarly, Waitzkin (1990) saw both qualitative and quantitative methods as facing fundamental problems and called for eclecticism in the choice of methods.

A similar perspective was developed by Walker (1985) specifically in relation to policy-oriented research. Interestingly, Walker described what he saw as the irreconcilable (at least at the current state of understanding) differences between the philosophical traditions which underlie qualitative and quantitative research.[4] However, in the face of the need to base policy on research, he advocated side-stepping such difficulties and making a choice between methods on purely pragmatic and instrumental grounds. He summarised his position as: "qualitative research reaches parts that other techniques don't" (18).

His defence of qualitative research methods lay in the inadequacies of more quantitative approaches: "on occasions when the more established research techniques are impotent, qualitative methods can yield information of considerable value to decision makers and social scientists" (3–4).

Walker saw a potential role for qualitative research at all stages of policy-related research, but he suggested that their flexibility and interactive potential is particularly important in the preliminary phases of research.[5] The decision about whether or not to use qualitative methods in a main study was seen to be closely related to the research topic. Where the topic is sensitive, complicated, poses measurement problems, or is concerned with process and/or interaction, qualitative methods may be the methods of choice. Similarly, qualitative methods may be indicated where the research subjects are either inarticulate or of high status. If the research population is very small or difficult to locate, qualitative methods may again be preferred.

Walker also suggested circumstances in which qualitative research might helpfully be used to complement quantitative methods. Specifically qualitative methods may be used to describe the process of an intervention.[6] while quantitative methods are used to measure the impact. On the other hand, qualitative methods may be used to interpret, illuminate, illustrate or qualify quantitative findings.

Having outlined the differences in approach between experiments and qualitative research, LeCompte and Goetz (1982) argued that the sharing of data collection strategies from the two approaches is perfectly legitimate. Qualitative research may serve to augment the reliability or validity of an experimental design, by providing a procedural and contextual frame for experimental manipulation. Similarly, in quantitative research, researchers may use deliberate manipulation to "elicit participant sanctions for the violation of social norms or to provoke other reactions from the subjects of a study" (35).

Specifically, in the context of health promotion, Secker and co-workers (1995) advocated a 'horses for courses' approach to the choice of paradigm,[7] depending upon the object of the research. Rather than seeing the choice of methods as dictated by questions of ontology and epistemology,[8] they argued that the researcher should choose his or her philosophical position on the basis of the task in hand. Thus, they suggested, when one is carrying out developmental research, a relativist perspective is appropriate. On the other hand, a more empiricist, realist perspective is appropriate when the aim is description of overt opinions, testing elements of health promotion programmes.[9]

3.1.1.3 Qualitative research as senior partner

A third perspective within this broadly pragmatic and instrumental approach to the use of qualitative methods can be discerned. Here a hierarchy of methods is implied, but in this case qualitative methods are at the top of the hierarchy and alternative methods are to be measured against these. An

[4] These alleged philosophical differences are considered in section 3.2.1; [5] See section 3.2.2.5; [6] See section 3.2.2.4; [7] See section 3.1.2 for a discussion of paradigms; [8] See section 3.2.1; [9] See section 3.2.1.1. Subtle realism is an alternative to radical realism and radical relativism.

early version of this position was argued in a classic paper by Becker and Geer (1969a), in which they argued for the superiority of the method of participant observation,[10] on the grounds that it provides "the most complete form of sociological datum" (322). They accepted that there might be good pragmatic reasons for using other qualitative and quantitative methods and which might include practicality, economy and issues of research design. However, such alternative methods were seen as second best.

Becker and Geer's original paper provoked critical responses from some of their contemporaries (e.g. Trow, 1969). In reaction to these, they either modified or clarified their position (Becker and Geer, 1969b). They insisted that their argument simply referred to the completeness of data obtained by various methods and that such completeness was only one of the criteria which should properly be applied to the evaluation of different research methods. Other criteria might include relevance, accuracy and reproducibility.

In the same rejoinder, Becker and Geer made explicit their adherence to the instrumental camp in the debate about the relative merits of qualitative and quantitative methods:

> We do not argue that participant observation should be used in all studies, but simply that it is possible to tell by comparison with the data it produces what data is lost by the use of another method. Whether the loss is important or not depends upon the character of the problem under investigation; whether the loss is unavoidable or too expensive is a practical, not a logical problem. (340)

Finch (1986) also adopted this approach of qualitative research as senior partner, at least in relation to some aspects of research. She argued that qualitative research has particular strengths in the study of the process and context of social action[11] and in reflecting the subjective reality of those being studied.[12]

Many of the advocates of the broad instrumental perspective on the place of qualitative methods, appear to see the purpose of research, whether in the social or natural sciences, as the production of a true reproduction of the world.[13] For example, Vidich and Shapiro (1969) were concerned to compare survey and qualitative data in terms of their validity, where validity is defined as, "a correct account of [a group's] attitudes, values and practices" (28). Similarly Becker and Geer (1969a) advocated the use of participant observation on the grounds that it allows the researcher to check description against fact. Thus it is capable of exposing unintentional misrepresentation of the true state of affairs by informants: "Any such mythology [the belief that superiors in an organisation are out to get you] will distort people's views of events to such a degree that they will report as facts things which have not occurred, but which seem to them to have occurred" (328).

Walker (1985) recognised the problems, which are raised by such naive realist assumptions.[14] However, his commitment to pragmatism in the service of policy relevant research leads him to at least act as if realist assumptions are justified. Hammersley (1992d) argued that a modified form of realism, subtle realism avoids some of the problems associated with naive realism and is compatible with instrumentalism.[15]

This assumption that the purpose of research is to develop true accounts, whether they are descriptions or explanations, of the way the world is, dictates the criteria by which such accounts are to be judged. The question of which methods from the 'tool-kit' are most appropriate in a given situation is translated into a question about the extent to which a given method decreases the likelihood of an untrue or invalid account of its object.[16]

A secondary criterion, that of efficiency, may also be applied. Whatever the evaluator's conclusion about the method which is, in principle, most appropriate for answering a particular question, if there were no constraints of time or money, it is nevertheless reasonable to ask what, in the real world, represents the best trade-off between validity and efficiency.

3.1.2 The two-paradigms approach

Conversely, there are those who argue that qualitative and quantitative research methods are best seen as deriving from fundamentally different paradigms (e.g. Smith, 1985; Smith and Heshusius, 1986; Dootson, 1995). Here the notion of a paradigm is taken from the work of Kuhn (1962). Guba and Lincoln (1994) defined a paradigm as "the basic belief system or world view that guides the investigator, not only in choices of method but

[10] See section 4.2; [11] See sections 3.2.2.3 and 3.2.2.4; [12] See section 3.2.2.1; [13] A critique of the reproduction model of research can be found in section 3.2.2.2; [14] See section 3.2.1.1 for further discussion of the realist/relativist debate and the problems of naive realism; [15] Subtle realism is discussed in section 3.2.1.1; [16] See chapter 5.

in ontologically and epistemologically fundamental ways" (105).

These beliefs are treated as basic in the sense that they have to be considered as fundamental axioms which are accepted on faith, however well argued they may be. There is no possibility of establishing their essential truthfulness.

The focus here is upon the philosophical assumptions which underpin research enquiries and relate to both ontological and epistemological concerns.[17] In particular, much of the discussion centres around the nature of social reality, how it should be studied and the question of what is to count as warrantable knowledge. Central to this position is the assumption that the distinction between qualitative and quantitative techniques involves much more than a debate about quantification. This was summed up by Rist (1977 cited in Bryman, 1988):

> When we speak of 'quantitative' or 'qualitative' methodologies we are in the final analysis speaking of an interrelated set of assumptions about the social world which are philosophical, ideological and epistemological. They encompass more than simply data gathering techniques. (67)

Similarly Lincoln and Guba (1985) and Lincoln (1990) have argued that paradigms are pervasive and ineluctable:

> The adoption of a paradigm literally permeates every act even tangentially associated with inquiry, such that any consideration even remotely attached to inquiry processes demands re-thinking to bring decisions into line with the world view embodied in the paradigm itself. The immediate realization is that accommodation between paradigms is impossible. The rules for action, for process, for discourse, for what is considered knowledge and truth, are so vastly different that, although procedurally we may appear to be undertaking the same search, in fact, we are led to vastly diverse, disparate, distinctive and typically antithetical ends. (Lincoln, 1990:81.)

The pervasive nature of different paradigms is such, Lincoln argued, that attempts at a mix-and-match strategy are tantamount to courting disaster. She pointed to an intra-psychic need for coherence, order and logic, which demands that inquirers make a commitment to one or other paradigm, and added: "To do otherwise is not only to commit paradigmatic perjury, it is to invite psychological disaster" (81).

While Lincoln identified qualitative research as central to the constructivist paradigm, she espouses, this is only one element of a much broader programme, the central characteristic of which is a commitment to a relativist ontology.[18] Indeed Guba and Lincoln (1994) argued for the primacy of paradigm over method. They held that, in principle, any paradigm may incorporate qualitative and quantitative methods. Nevertheless, it seems clear that they see qualitative methods as being fundamentally more suited to the constructivist paradigm which they adopt.

Smith and Heshusius (1986) argued that the recent history of qualitative research can be seen as a movement away from this position, which emphasised the fundamental, paradigmatic incompatibility of qualitative and quantitative methods, to one that sees qualitative and quantitative methods as compatible and which calls for cooperation between practitioners of the two approaches. They view this rapprochement as the premature closing down of an important conversation, which is the result of the failure to distinguish between two different definitions of method. Where method is interpreted merely in terms of research technique, mixing methods may be possible. Where, as Smith and Heshusius claim it must be, method is concerned with more fundamental questions about the logic of justification, compatibility can only be achieved where one tradition abandons its own logic of justification and adopts criteria derived from the other. It is Smith and Heshusius' contention that, in moving closer to quantitative methods, qualitative researchers have abandoned the anti-realist epistemological foundations upon which qualitative research is based.[19]

The polarising tendency of this approach of two incommensurable paradigms has been subjected to extensive critique, even among those who are sympathetic to qualitative research (Reichardt and Cook, 1979; LeCompte and Preissle, 1993a). In section 3.2, we shall consider critically the dimensions of the inter-paradigmatic differences, which are proposed by advocates of this position.

3.1.3 Critical theorists[20]

Third, there are those for whom the choice between qualitative and quantitative research is best made upon ideological and/or political

[17] See section 3.2.1; [18] See section 3.2.2.1; [19] See section 3.2.1.1; [20] See section 2.3.1 for a discussion of the historical background to this position in the work of Kant.

grounds. Guba and Lincoln (1994) used **critical theorists** as a blanket term to denote a range of positions which insist upon the value-determined nature of inquiry. This term is taken to include neoMarxism, feminism, materialism and participatory inquiry. What can be known is seen as linked inextricably with the interaction between a particular investigator and particular research participants. The implications of this position for methodology are that inquiries are understood as dialogues between the researcher and the researched, which are dialectical in nature insofar as they seek to transform ignorance and misapprehensions into more informed consciousness.

This position has been developed most fully by feminist methodologists who typically argue that quantitative research is incompatible with feminism insofar as it involves the exploitation of already exploited women (Oakley, 1981; Stanley and Wise, 1983). It is also adopted by those who advocate more participative approaches to research, including those seeking to empower ethnic minority groups such as Australian aboriginals (Baum, 1995).

The methods advocated by critical theorists are predicated upon their commitment to illuminate the impact of the distribution of power, privileges, resources, status, authority, leadership and decision making (LeCompte and Preissle, 1993a). In general, such standpoint research rejects traditional science with its emphasis, as they see it, on control, omission of context and objectification. Both postmodernists and post-structuralists can be seen as related to this critical theorist perspective. In particular both reject existing authority and hence the notion of **objective** social science inquiry (LeCompte and Preissle, 1993a).

3.2 The dimensions of the qualitative/quantitative debate[21]

Having discussed the range of views which are proposed by researchers and methodologists about the appropriate relationships between qualitative and quantitative methods, we now move on to consider some of the terms in which the qualitative/quantitative debate has been conducted. In keeping with the two-paradigms approach outlined above, it has become conventional to argue that qualitative research and quantitative research represent opposite sides of a number of different polarities. Finch (1986) has suggested that this representation of qualitative and quantitative research as antithetical reflects the reality that the qualitative research tradition has been primarily oppositional, insofar as, conventionally, it has been seen as posing a challenge to the dominant (quantitative) research tradition. Particularly since World War II, quantitative research has achieved an ascendant position, which means that those who do not conform to its methodological imperatives are called upon to justify their position. Such justifications have often taken the form of critiques of the assumptions which are thought to underlie quantitative methods, rather than claims about qualitative methods *per se*. As Silverman (1989) has observed, qualitative researchers are generally more adept at criticising quantitative methods than at developing their own programme.

This polarising approach has been challenged, particularly by those who believe that the choice of methods should be based upon instrumental and pragmatic considerations (Reichardt and Cook, 1979; Bryman, 1988; Hammersley, 1992a; LeCompte and Preissle, 1993a; Silverman, 1993).[22] These authors suggested that such readiness to place qualitative research on opposite sides of a set of polarities is misleading on a number of grounds, not least that such polarities seriously misrepresent both qualitative and quantitative research as they are currently practised. Qualitative and quantitative research are seen as being much more heterogeneous than the dichotomising approach allows, so that some qualitative research adopts the assumptions and practices that are attributed to quantitative research and vice versa. Where there are differences in the practices of qualitative and quantitative researchers, these are better seen as representing different emphases (Atkinson, 1979). The implication of this position is that the choice of method is best made on the basis of the purposes and circumstances of research (Hammersley, 1992g; Silverman, 1993), rather than ideological or methodological commitment.

In the rest of this section we shall consider a number of the dimensions along which it is sometimes argued that qualitative and quantitative research differ fundamentally. We shall consider the dichotomies which are proposed, the arguments which are presented in support of such dichotomies and we shall discuss whether these dichotomies represent necessary, or indeed actual differences between the commitments and

[21] See section 2.3 for historical background to this debate; [22] See section 3.1.1.

practices of qualitative and quantitative researchers. We shall first consider some of the philosophical dimensions of the debate (section 3.2.1), before considering some of the characteristics of research practice which are sometimes seen as definitive of the qualitative tradition (section 3.2.2.).

3.2.1 Philosophical considerations and assumptions

Whether or not the philosophical bases of qualitative and quantitative research are judged to be antithetical, it is important for researchers working in both traditions to be clear about the philosophical assumptions that they are making as they design their research and draw conclusions from it. As Hammersley (1992d) argued, while philosophy is not foundational, in the sense that it is not necessary to solve all philosophical problems before embarking upon social research, nevertheless:

> There is no escape from philosophical assumptions for researchers. Whether we like it or not, and whether we are aware of them or not, we cannot avoid such assumptions. (43)

3.2.1.1 Idealism versus realism[23]

The two-paradigms approach[24] is consistently argued by Smith (Smith, 1983a; Smith, 1983b; Smith, 1984; Smith, 1985; Smith and Heshusius, 1986; Smith, 1989). He rejected any attempt to treat qualitative and quantitative approaches as complementary or to base the choice of qualitative or quantitative methods as driven by the nature of the problem to be investigated, or, as Silverman (1993) has put it "what you are trying to do" (22). For Smith, there is a fundamental ontological cleavage which separates qualitative and quantitative research, such that the former is inextricably grounded in scientific idealist assumptions, while the latter research is based upon scientific realism (see also Guba, 1990).

The historical background to these two philosophical positions has been described in chapter 2 of this report. Smith (1983a; 1983b) traced realism back through Comte, Mill and Durkheim to Newton and Locke. Idealism, on the other hand, is associated with Dilthey, Rickert and Weber. As Smith (1983a; 1983b) pointed out, we should be cautious of treating scientific idealism as identical to the position adopted by Dilthey or Weber. Similarly, scientific realism cannot be equated with nineteenth century positivism[25] and indeed one very influential version of positivism (logical positivism) was itself anti-realist (Williams and May, 1996). These authors suggested that the empiricism of Hume and Locke, with its insistence that all that is accessible to us is the world of appearances, had overtones of philosophical idealism. The complex relationship between realism and positivism was discussed in some detail by Blaikie (1993). Nevertheless, there are important continuities as well as discontinuities between nineteenth century thought and the contemporary debate, which are best understood against the background of the material presented in chapter 2. In this section we shall summarise the two broad positions and consider their implications for the practice of social research in relation to HTA.

Williams and May (1996) defined realism as the belief that "the world has an existence independent of our perception of it" (81). They quoted Bhaskar's definition: "Normally to be a realist in philosophy is to be committed to the existence of some disputed kind of being (e.g. material objects, universals, causal laws, propositions, numbers, probabilities; efficacious reasons, social structures, moral facts)" (Bhaskar, 1993:308 cited in Williams and May, 1996:81).

Smith (1983a; 1983b) summed up the position in more colloquial language:

> Realism is based on the idea that reality exists independent of us. Independent means that this reality exists whether or not we are aware of it or take any interest in it. (Smith, 1983b:8.)

In Heap's terms realism is therefore dualist, rather than monist (Heap, 1995). In particular, realists hold that the entities described by theories, whether in the natural or the social sciences, really do exist (Bhaskar, 1975; 1979). The claims made by theories are either true or false. The object of science is to establish the truth about how the world operates. This applies to the social world as well as to the natural world (Blaikie, 1993).

Scientific idealism (the philosophical position which Smith (1983a; 1983b; 1985; 1989) believes to be fundamental to qualitative research) can be summarised as the view that the external world consists merely of representations and is a creation of the mind (Williams and May, 1996). Thus, it is a **monist** position (Heap, 1995). Idealists do not necessarily deny the possibility of material

[23] See sections 2.2 and 2.2.3 for an historical perspective on this debate; [24] See section 3.1.2; [25] See section 2.3.2 for a discussion of nineteenth century positivism.

existence. Rather it is the knowability of any such reality which is the central issue. Williams and May (1996) took an example from the physical world to illustrate the idealist position:

> Next time you are in a room containing a table, or a desk, look at it from above and note its descriptive characteristics. Now get on your hands and knees and look at it from underneath, now look at it sideways on. Does it look very different? Which was the 'real' table? Each of the tables you perceived was the same one, but if the experiences had been separated you could not have known this. Can we ever know the real table? This argument can be extended into the social world. However much we 'carve up' social interactions or social structure, we can never claim to have found out what is 'real' about it. ... In contrast to this we have a series of representations. (70–1)

Idealists reject the possibility of adopting a "God's eye point of view" (Smith, 1984:381). Rather than assuming that there is one reality, which the investigator must seek to track down, idealists hold that there are multiple realities, or as Smith put it, "as many realities as there are persons" (Smith, 1984:386). Smith (1983a; 1983b) distinguished between two versions of scientific idealism. The stronger version (ontological idealism) holds that social and human reality are **mind-created**. The weaker version (conceptual idealism) holds that reality is **mind-shaped**. In both cases there is a refusal to treat reality as existing 'out there'.

Guba and Lincoln (1989) adopted a broadly similar position to Smith's, though they preferred to use the term **constructivist** to describe their position. This position holds as axiomatic the idea that: "There exist multiple, socially constructed realities ungoverned by laws, natural or otherwise...these constructions are devised by individuals as they attempt to make sense of their experiences... constructions can be and usually are shared...this does not make them more real but simply more **commonly** assented to" (86, original emphasis).

This constructivist position involves a rejection of the notion that the world is composed of facts and that the role of research is to uncover those facts (Palmer, 1928). According to Schwandt (1997), the central idea of constructivism is that what we take to be objective knowledge or truth is irreducibly the result of perspective. In that sense, what we take to be facts are inescapably theory-laden (Guba and Lincoln, 1994). The immaculate perception of untheory-laden facts available in the same way to every observer' is rejected.[26]

Hammersley (1992d) related such anti-realist trends within social science to developments within the philosophy of science, associated with the work of Kuhn (1962) and Feyerabend (1975). The notion that science operated by accumulating secure and precise knowledge based on observation and logic came under challenge. For Kuhn, science proceeds through periods of paradigm consensus. In each such period, scientists operate with a set of largely uncontested assumptions about the phenomena to be investigated and the methods by which they are to be understood. These periods of consensus are punctuated by scientific revolutions, when such uncontested assumptions are challenged and overthrown. Such revolutions do not arise simply from a rational assessment of empirical evidence. Hammersley (1992d) suggested that this debate about the status of scientific knowledge had a significant influence on the rise of anti-realism within the social sciences. This is reflected in the enthusiasm with which some idealists/relativists embrace the language of incommensurable paradigms in discussing the philosophical foundations of qualitative and quantitative research traditions.

Schwandt (1997) distinguished between two forms of constructivism. The first, **radical constructivism**, involves a denial that it is possible to have any knowledge of a phenomenon apart from our own experience of it. The second, **social constructionism** focuses upon knowledge as inter-subjectively rather than individually constructed. Within this second version, the accounts which individuals give of the world are not treated as the expression of individualised, internal processes. Rather they are understood as an expression of the relationships among persons. (See Heap, 1995 for a similar distinction between cognitive and social constructionism.)

The implications of adopting a realist or idealist ontology for the research enterprise become clear when one considers the relationship between these ontologies and the theories of truth which they imply. Thus ontological debates raise epistemological debates. Hammersley (1992d) made the link between realism and correspondence theories of truth, when he defined realism as "the idea that there is a reality independent of the researcher

[26] See section 2.3.1 for historical background to this discussion of the theory-driven nature of observation in Kant's writings.

whose nature can be known and that the aim of research is to produce accounts that correspond to that reality" (43).

Such correspondence theories hold that truth is the degree to which an account corresponds to a phenomenon which it claims to represent. As Hammersley (1993) discussed, there are some serious difficulties associated with such correspondence theories. The most fundamental of these is the so-called **problem of the criterion**. This was discussed in some detail in chapter 2 in relation to the post-sceptics and hyperbolic doubt.[27] The problem is that of finding any dependable grounds for making a truth claim. Hammersley summed up the problem:

> Faced with someone making a claim to knowledge, the sceptic asks what the grounds are for the claim, and when offered reasons for believing the claim to be true, inquires what the reasons are for believing those reasons, and so on. In this child-like manner sceptics seek to demonstrate that the provision of evidence either goes on for ever or is circular, appealing to something that is itself part of what is to be established. (17)

The problem is one of infinite regress. There is no basis, which is beyond doubt, upon which we can base any claim to certainty. In practical terms, since there is no way of checking the correspondence between any account of a phenomenon and reality, there is no possibility of establishing the truth or falsity of any account.

There have been a range of responses to this problem of the criterion. Some have attempted to save correspondence theories of truth by identifying some body of knowledge of whose truth we can be absolutely certain, and using this as a secure foundation upon which to build further knowledge. One version of such **foundationalism**, that of Descartes, was discussed in chapter 2.[28] A more recent example is the attempt by positivist foundationalists to treat basic sense data as the only basis upon which scientific study may proceed. Kolakowski (1993) summarised this position: "We are entitled only to record that which is actually manifested in experience; opinions concerning occult entities of which experienced things are supposedly manifestations are untrustworthy" (3).

Unfortunately, as Hammersley pointed out, this reliance upon experience as the foundation for knowledge is not sufficient to save correspondence theories of truth. As Hammersley (1993) argued, reliance upon basic sense data does not free us from the challenge that these sense data may, both in principle and in practice, mislead us. Optical illusions are just one example of the way in which what we take to be the evidence of our senses is capable of misleading us. This is the same problem which preoccupied the post-sceptics.[29]

Idealism can be seen as an alternative response to the problem of the criterion. Rather than seeking to identify an incontrovertible foundation for knowledge, idealism side-steps the problem by rejecting correspondence with reality as an appropriate definition of truth. Truth is seen as context-bound and it is possible for multiple truths to exist. Smith (1984) identified two alternative positions on truth which relate to the two forms of idealism (conceptual idealism and ontological idealism) outlined above. Conceptual idealists, who hold that reality is **shaped** by mind, espouse a coherence theory of truth. Williams and May (1996) suggested that coherence is a stronger relationship than mere lack of disagreement. It involves inter-subjective agreement. A statement is accepted as true if there have been numerous reports which are coherent with one another and there have been no reports to the contrary. However, they raised significant problems with such coherence theories of truth. They pointed out that it is possible for statements to be coherent without being true. They took the example of reports that koalas are bears. Despite the coherence of multiple reports that koalas are bears, such reports are untrue, as koalas are marsupials and not bears, though this classification is itself the result of a social process leading to a consensus about relevant taxonomic criteria. Such reports are highly coherent but wrong.

Ontological idealists, who hold that reality is mind-created, rather than mind-shaped, also reject correspondence theories of truth. They hold to an even weaker version of truth than conceptual idealists. For them, truth is no more than socially or historically conditioned agreement. Guba and Lincoln (1989) elaborated this position, by defining truth as "that most informed and sophisticated construction on which there is consensus among individuals most competent (not necessarily most powerful) to form such a construction" (86).

Here there is no commitment to uncovering the ultimate truth. For constructivists such as Guba and

[27] See section 2.2.2; [28] See section 2.2.2; [29] See section 1.2.2.

Lincoln it is possible for multiple and even contradictory truths to exist side by side and any of these truths may be overthrown at any time, should a "really disruptive insight come to light" (87).[30] Similarly, Smith (1984) defined truth as "socially and historically conditioned agreement".

Ontological idealism, in particular, poses substantial problems for those who wish to establish a body of knowledge over which there is inter-subjective agreement (Hammersley, 1992d; Atkinson and Hammersley, 1994; Heap, 1995; Campbell, 1994; Greene, 1996; Williams and May, 1996; Schwandt, 1997). If knowledge is treated as mind-dependent (Williams and May, 1996), the concept of truth becomes somewhat irrelevant. Indeed, Lincoln and Guba themselves acknowledged this when they argued that: "It is dubious that the constructivist paradigm requires a term like **truth**, which has a final or ultimate ring to it." (Guba and Lincoln, 1989:86). Thus idealism is often associated with relativism or the denial that there can be any ultimate truths. The researcher's account of the world comes to be seen as just one version of the world among others.

A number of authors have identified what they see as profoundly negative consequences of adopting a radically relativist position, such as that outlined by Guba and Lincoln. In critiquing the application of Guba and Lincoln's position to evaluation research, Greene (1996) observed that, from such a position "one evaluator's understandings about the experiences and impacts of a given social program will be different from another evaluator's understanding, because such understandings are constructed by each inquirer's unique interactions with others in that context" (279).

Greene reviewed the conventional critiques of such relativism, which include the logical inconsistency of this position (briefly that, if everything is relative, then even the statement that 'everything is relative' is relative and therefore no more warrantable than the statement that 'everything is not relative'), that it encourages the chaos of everything goes (279) and that it undermines the usefulness of research findings for practice. These are clearly significant concerns in relation to the practice of HTA. If researchers adopt the radical relativist position outlined above, then it is difficult to see how health service commissioners could make use of their findings.

Similarly, Hammersley (1992d) described the fundamental problems which the acceptance of this relativist position raises for social researchers:

> If it is true that what ethnographers produce is simply one version of the world, true (at best) only in its own terms, what value can it have? And there is no reason to suppose ethnographers produce just one version of the world. Given that they differ among themselves in cultural assumptions, we must surely conclude that their accounts are to be viewed as creating multiple, incommensurable worlds on the basis of the same or similar research experience. ... If this is so, what is the point of spawning yet more versions of 'reality' given the relative costs of ethnography compared with, say, armchair reflection? And why should some 'realities' be published and discussed at the expense of others? (49)

Or why, we might ask, should the NHS Executive sponsor research which has no greater claim to dependability than that it creates just one among the multiple, incommensurable versions of reality which are possible? Heap (1995) argued that a thorough-going ontological idealism is fundamentally incompatible with policy-oriented or evaluation research.

Again, Schwandt (1997) acknowledged that one of the problems associated with research which adopts an idealist/relativist/constructionist stance is that it lacks any critical purchase on the accounts which members of a social setting give of that setting. Given the assertion of multiple, and potentially competing, realities, there is no possibility of critiquing such accounts. Members' accounts are, in effect, treated as incorrigible.

Atkinson and Hammersley (1994) expressed concern that if researchers adopt a consistently sceptical and relativist stance, the end result is likely to be what they describe as a **debilitating nihilism** (252). Campbell (1994) also believed that idealist/relativist positions, when pushed to their logical conclusions, imply an abandonment of both the effort to know and the possibility of "mutual criticism as to the validity of such efforts" (154).

The choice between a realist or idealist position also has important implications for the relationship between the researcher and what (s)he is investigating. Smith (1983b) argued that a realist position implies that it is possible to effect a separation between the observer and what is observed. The

[30] Heap (1995) argues that while Lincoln and Guba adopt the rhetoric of monist ontological idealism, they advocate research and evaluation practices which are inconsistent with this position.

researcher's activity is directed towards discovering an independently existing reality. The researcher is in a subject–object relationship with what is being researched. The implication of this, he argued, is that the social scientist is in the same relationship with the object of his research as is the physical scientist. This is not to imply that the concepts of social research will necessarily have the same material existence as some of those studied by physical scientists. Rather they have the same status as non-observable entities such as electrons. Bhaskar (1975; 1979) used the example of magnetic fields to make the same point. While magnetic fields cannot be directly observed, we can know them through their causal effects. Similarly, he argued, the mechanisms at work in the social world can be discovered through observation of their causal effects.

By contrast, Smith (1983b) argued, the idealist is in a subject–subject relationship with what is being researched, insofar as the latter is understood as depending upon the "constituting activities of our minds" (8). Smith saw the implication of adopting an idealist position as being that one cannot separate the process of investigation from what is being investigated, in the way which is often seen as crucial to scientific investigation: "The mind involvement of a constituted reality, and hence the impossibility of its existence as an independent reality, means that the process of investigation itself will affect what is being investigated" (9).

Smith also discussed the implications of realism and idealism for the status that is given to research instruments such as intelligence tests. Whereas, for realists, they are an attempt to measure some independently existing reality, for idealists such measures are one element in the process of constituting intelligence. The same arguments would, of course, apply to research instruments designed to measure, for example, psychiatric morbidity, locus of control or neuroticism.

As Guba and Lincoln (1989) pointed out, the choice between an idealist and a realist position cannot be made on the basis of fact. Rather, they are based upon *a priori* assumptions which are not open to investigation or refutation using the tools of science. However, for the purposes of this review, it will be helpful to consider two questions. First, to what extent are authors such as Smith correct to claim that qualitative research is inextricably associated with idealist assumptions, while quantitative research is bound up with realism? Second, do the problems associated with correspondence theories of truth, particularly the problem of the criterion discussed above, pose insuperable problems for those who adopt a realist position.

Hammersley (1992d) raised serious doubts about the extent to which qualitative research is, in practice, identified with idealist assumptions. Indeed, he argued, many ethnographers have justified their use of naturalistic methods of data collection in terms of the particular strengths of such methods in gaining access to the true nature of social phenomena. He cited Blumer's use of metaphors such as 'lifting the veils' and 'digging deeper' as evidence of the realist assumptions underlying the ethnographic enterprise. Bryman (1988) made a similar point about the empiricism which he identified in the work of many qualitative researchers, in particular sociologists of deviance, such as Douglas. However, as Hammersley (1992d) recognised, there are significant counter-pressures to this realist commitment within ethnography. He argued that, while ethnographers, such as Blumer, have adopted a realist ontology in respect of their own investigative activities, they have tended to adopt an idealist ontology in respect of the people they seek to investigate: "Central to the way in which ethnographers think about human social action is the idea that people **construct** the social world, both through their interpretations of it and through the actions based upon those interpretations" (44).

There is, for Hammersley, a fundamental contradiction in treating the beliefs and perspectives of the social actors we are investigating as incommensurable, socially constructed versions, while treating the accounts which ethnographers produce as more or less accurate representations of some externally existing reality. Hammersley and Atkinson (1995) described this as a tension within ethnography between the naturalism characteristic of ethnographers' methodological thinking and the constructivism and cultural relativism that shape their understanding of the perspectives and behaviour of the people they study (11). Similarly, Schwandt (1997) referred to the paradox of qualitative research: researchers struggle to develop an objective interpretive science of subjective human experience.

Elsewhere, Hammersley (1992g) pointed out that not all quantitative researchers are realists. For example, he cited Lundberg, an influential advocate of quantitative method in the social sciences, as arguing that the "only universe with which science can deal is 'our conception' of it" (Lundberg, 1933:309 cited in Hammersley, 1992g:171). Conversely, Harre's advocacy of

qualitative method has been linked with an explicit commitment to realism (Harre, 1970).

We have seen that, in practice, some qualitative researchers do ground their work in realist assumptions, while, conversely, some quantitative researchers draw upon idealism. However, this still leaves us the question of whether, regardless of the practice of individual researchers, there is any necessary relationship between qualitative methods and particular ontological or epistemological positions as, for example, Smith (1983a; 1983b) has argued.

Hammersley (1992d) asserted that, in the naive versions, discussed above, neither idealism nor realism offer a sound philosophical basis for social research. As already discussed, naive relativism renders social research pointless. Naive realism fails to recognise the sheer impossibility of unmediated contact between the investigator and the object of his or her investigations. The risk of naive realism, as Campbell (1994) pointed out, is that researchers fail to recognise that their perceptions are a "joint product of the referent and the cultural–biological lenses through which it [the phenomenon of interest] is seen" (157). As a result, the observations may be treated inappropriately as wholly objective. Campbell emphasised the importance of recognising the "powerful effects of culture on cognition" (157).

Hammersley (1992d) argued for a third position, which he calls **subtle realism**. In adopting this position, the investigator concedes that it is impossible to have certainty about any knowledge claims. In that sense, his position is non-foundationalist (Hammersley, 1993). Similarly, the subtle realist accepts that there is no way in which the researcher can escape the social world in order to study it (Hammersley and Atkinson, 1995:17). Rather the objective should be the search for knowledge about which we can be reasonably confident. Such confidence will be based upon judgements about the credibility and plausibility of knowledge claims.[31] Subtle realism maintains that phenomena exist independently of the investigator's claims about them and that the investigator's claims may be more or less accurate. In claiming that phenomena exist independently of the investigator, Hammersley is saying that any claim which the investigator makes about reality does not, in itself, change the nature of those phenomena in such a way as to make the claim either true or false. Finally, and most importantly, subtle realists see the aim of social research representing reality, rather than reproducing it. This is an important distinction between subtle and naive realism. For the subtle realist, any given reality can be represented from a range of different perspectives. Each of these representations may be treated as true. This approach accepts that representations of reality are always representations from a particular point of view and that it is futile to search for a "body of data uncontaminated by the researcher" (Hammersley and Atkinson, 1995:16). This opens up the possibility of multiple, **non-competing**, valid descriptions and explanations of the same phenomenon. However, it excludes the possibility of multiple, **competing**, valid descriptions or explanations of the same phenomena.

Like Hammersley, Phillips (1990) accepted the impossibility of making theory-free observations of brute facts. However, just because observations are necessarily theory-laden does not necessarily mean that it is impossible to judge between them. Phillips argued that it is possible to combine a commitment to the social construction of reality with a concern for truth as a regulatory ideal, through studying the different constructions people make of reality, without accepting that particular beliefs are true.

A similar point was made by Bhaskar (1979) when he argued for a careful distinction between **epistemic relativity**, which he endorsed, and **judgemental relativism**, which he rejected. The former, epistemic relativity, "asserts that all beliefs are socially produced, so that all knowledge is transient, and neither truth-values nor criteria of rationality exist outside historical time" (73). The latter, judgmental relativism, "asserts that all beliefs (statements) are equally valid, in the sense that there can be no (rational) grounds for preferring one to another".

Hammersley (1992d) argued that subtle realism offers a middle position between naive realism and naive idealism, freeing us from the necessity, posited by Smith, (1983a; 1983b), to opt for one or other polarity. This middle-way allows us to accommodate some elements of social constructivism, without abandoning a commitment to independent truth as a regulative ideal:

> This subtle realism retains from naive realism the idea that research investigates independent, knowable phenomena. But it breaks with it in denying that we have direct access to those phenomena, in accepting

[31] This is more fully discussed in chapter 5.

> that we must always rely on cultural assumptions, and in denying that our aim is to reproduce social phenomena in some way that is uniquely appropriate to them. Obversely, subtle realism shares with scepticism and relativism a recognition that all knowledge is based on assumptions and purposes and is a human construction, but it rejects these positions' abandonment of the regulative ideal of independent and knowable phenomena. Perhaps most important of all, subtle realism is distinct from both naive realism and relativism in its rejection of the notion that knowledge must be defined as beliefs whose validity is known with certainty. (Hammersley, 1992d:52.)

Hammersley (1992d) presented this subtle form of realism as equally appropriate for qualitative and quantitative social research, thus challenging Smith's assertion that qualitative and quantitative research are linked to fundamentally different philosophical paradigms. It can be seen as an alternative to the ontological and epistemological dichotomies proposed by some qualitative researchers which, we suggest, represent neither an accurate analysis of the current state of qualitative and quantitative research, nor a necessary or helpful distinction between these two traditions.

3.2.1.2 Induction versus deduction[32]

In this section we are concerned with the logic of research and, in particular, the place of theory in the research process. As Hammersley (1992g) observed, qualitative research is often characterised as adopting an inductive research strategy as compared to quantitative research which is characterised as deductive (Shaffir *et al,* 1980; Finch, 1986; Imle and Atwood, 1988; Merriam, 1988; Stange and Zyzanski, 1989; Habermann-Little, 1991; Munhall, 1993b; Morse, 1994; Clarke, 1995; Lindlof, 1995). Induction and deduction are based on alternative types of argument (Blaikie, 1993). Research that is purely deductive begins with a theoretical system, operationalises the concepts of that system and then sets out to gather empirical data to test that system. Research that is purely inductive, on the other hand, starts with the collection of data and moves from there to a general conclusion. It involves the derivation of a general principle from a set of specific observations (Williams and May, 1996).[35]

Our examination of the claim that qualitative and quantitative research are characterised by different logical structures will fall into three parts. First, we shall consider the argument that the logic of quantitative research is characteristically deductive. Then we shall go on to examine the position that qualitative research is essentially inductive. Finally, we shall some examples of the combination of inductive and deductive reasoning in qualitative research.

Quantitative research = deductive research? The claim that quantitative research is based on a exclusively deductive logic is attributable, at least in part, to the **critical rationalism** of Popper (1959). Popper sought to distinguish science from pseudo-science. He advocated a falsificationist strategy which depended exclusively upon deduction. Popper rejected induction as the basis for generalisations. Chalmers (1982) summarised Popper's position:

> Science starts with problems, problems associated with the explanation of the behaviour of some aspect of the world or universe. Falsifiable hypotheses are proposed by scientists as solutions to the problem. The conjectured hypotheses are then criticized and tested. Some will be quickly eliminated. Others might prove more successful. These must be subject to even more stringent criticism and testing. ... It can never be said of a theory that it is true, however well it has withstood rigorous tests, but it can hopefully be said that a current theory is superior to its predecessors in the sense that it is able to withstand tests that falsified those predecessors. (45)

Popper disassociated himself from any concern about the source of the hypotheses which it was the responsibility of scientists to seek to falsify. Scientists were to concern themselves exclusively with the logic of knowledge (Blaikie, 1993):

> The initial stage, the act of conceiving or inventing a theory, seems to me neither to call for logical analysis nor to be susceptible to it. The question of how it happens that a new idea occurs to a man – whether it is a musical theme, a dramatic conflict, or a scientific theory – may be of great interest to empirical psychology; but it is irrelevant to the logical analysis of scientific knowledge. (Popper, 1959:31–2.)

Similarly Reichenbach (1938) distinguished between the context of justification and the context of discovery. He argued that the context of discovery, the generation of theoretical ideas, lay outside science, while the context of justification was the realm of science.

However, such neglect of the source of the hypotheses which make up hypothetico-deductive theories has been criticised (Blaikie, 1993). Lincoln and Guba (1985) argued that it does not reflect

[32] See sections 2.2 and 2.3 for an historical perspective on debates about the practice and relative merits of deductive and inductive research; [33] See sections 2.2 and 2.3 for an historical background to this discussion of induction and deduction.

actual scientific practice, even in the natural sciences. They pointed out that much of Einstein's work would be excluded from such a definition of science. Wallace (1978) argued that the scientific process involves both induction, via empirical generalisations, and logical deduction leading to the testing of theories. The process is circular, rather than linear. Individual observations are synthesised into empirical generalisations, which, in turn can be synthesised into a theory via concept formation. This resulting theory may be transformed into new hypotheses through deduction and the resultant hypotheses can then be subjected to new observations, which in turn can be the subject of empirical generalisations, and so on. According to this influential version of the logic of scientific method, then, quantitative research is seen as involving both induction and deduction, rather than being exclusively deductive.

As Bryman (1988) and Hammersley (1992g) both have argued, it is also the case that, in practice, quantitative research is often highly inductive. Much quantitative research is concerned with description and empirical generalisation rather than hypothesis testing. In particular, factor analysis, which seeks to identify the underlying dimensions of a phenomenon on the basis of the clustering of data, is clearly an inductive technique.

Qualitative research = inductive? Next we turn to the assertion that qualitative research is primarily or exclusively inductive. Those who argue for this position often draw their inspiration from Glaser and Strauss (1967). It is certainly true that Glaser and Strauss were highly critical of what they saw as the obsession with theory testing which they believed characterised sociological research methods in the 1960's. Their critique of contemporary sociological research has much in common with Baconian scepticism, which was discussed in chapter 2.[34] They argued that this preoccupation with rigorous deductive theorising and testing had led to the neglect of the important task of theory generation. They opposed the imposition of *a priori* theory upon data and argued that, instead, the researcher should seek to generate **grounded theory**,[35] which, because it is derived from the data, rather than imposed upon it, will fit and work. Emerson (1983) defined grounded theory as theory which "grows out of, and is directly relevant to, activities occurring in the setting under study" (94). Glaser and Strauss (1967) described the process by which grounded theory emerges as inductive and this is, no doubt, one of the reasons why many qualitative researchers claim that qualitative research is, by definition, an inductive process.

However, closer inspection of the methods proposed by Glaser and Strauss show that grounded theory involves both induction and deduction (Schwandt, 1997). Grounded theory is discussed more fully in section 4.6. For the purposes of this section it is sufficient to observe that the constant comparative method of theory generation and refinement, proposed by Glaser and Strauss involves the constant movement between theory and data which is not dissimilar to that proposed by Wallace (see above). Indeed, Strauss and Corbin (1990) acknowledged that both deductive and inductive thinking are central to grounded theory analysis.

Lofland (1976) argued that the analysis of qualitative data is "simultaneously deductive and inductive" (66). In qualitative research, data analysis is not necessarily treated as a separate or discrete stage of research. Rather it occurs alongside data collection in such a way that it permits the researcher to follow up theoretical leads. Emerson (1983) suggested that it is, in fact, the movement backwards and forwards between theory and data, as theoretical statements are modified in the light of observations and observations are sought to extend or de-limit existing theory, that is characteristic of qualitative research. This combination of induction and deduction, which is sometimes termed **retroduction** (Bulmer, 1979; Hammersley, 1985; Blaikie, 1993) involves a double fitting of observations to theory and theory to observations (Emerson, 1983, citing Baldamus, 1972). Bulmer (1979) described the process:

> A theory is not pieced together from observed phenomena; it is rather what makes it possible to observe phenomena as being of a certain sort, and related to other phenomena. Theories put phenomena into systems. They are built up 'in reverse' – retroductively. (660)

The assumption that qualitative research is inductive rather than deductive is often translated into a claim that it is exploratory or hypothesis-generating rather than hypothesis-testing (Aamodt, 1983; Charmaz, 1983a; Duffy, 1985; Sandelowski, 1986; Imle and Atwood, 1988; Merriam, 1988; Stange and Zyzanski, 1989; Munhall, 1993b; Munhall, 1993c; Britten, 1995; Clarke, 1995). Again this at least partially reflects Glaser and

[34] See section 2.2.1; [35] See section 4.6.1.2.

Strauss's concern that the researcher should go into the setting being studied as close to *tabula rasa* as possible. However, as Emerson (1983) observed, other qualitative researchers take a less extreme position, arguing that the key issue is that the researchers should avoid **imposing** prior theoretical categories and assumptions on the data and remain open to new theoretical possibilities.[36]

Induction and deduction in practice in qualitative research. It is certainly the case that qualitative research has particular strengths in terms of hypothesis generation. For example, Gantley and co-workers (1993)[37] successfully used qualitative methods to generate hypotheses to explain the low incidence of sudden infant death syndrome (SIDS) in the Asian population. However, it is inaccurate to suggest that all qualitative research is either exclusively or even primarily concerned with hypothesis generation. As discussed above, even where qualitative research is avowedly inductive, it will often involve deductive aspects. As Hammersley (1992g) has commented:

> We cannot but rely on constructing hypotheses, assessing them against experience and modifying them where necessary. This is true whether we engage in hypothesis testing in a formal, explicit narrow way that involves subjecting hypotheses to crucial tests; or whether we adopt a more informal and broader approach in which we sacrifice some of the sharpness of the test in order to allow more of our assumptions to be thrown open to challenge. (169)

However, increasingly, qualitative researchers are seeking to engage in explicitly hypothesis testing research. This reflects a concern with the cumulative nature of research and a commitment to testing the limits and boundaries of the hypotheses generated in previous work. For example, when Silverman and co-workers (1992)[38] set out to analyse the organisation and reception of advice-giving in human immunodeficiency virus (HIV) counselling sessions, they took as their starting point Heritage and Sefi's findings about the delivery and reception of advice in interactions between health visitors and mothers (Heritage and Sefi, 1992[39]) Heritage and Sefi had found that most advice sequences were initiated by the health visitor, rather than the mother. They identified four different forms of advice-giving, which were observable in the interactions between health visitors and their clients. They analysed the relationship between different forms and the verbal responses of mothers to such advice. They concluded that one form of advice-giving, which involved **stepwise entry**, was associated with less resistance and more uptake, as evidenced by the mothers' use of marked acknowledgements. Such stepwise entry involved the health visitor eliciting a problem indicative response from the client and seeking further specification of the problem, before offering advice.

Silverman and co-workers (1992) sought to test hypotheses derived from Heritage and Sefi's work in the context of HIV counselling sessions. Like Heritage and Sefi they found a clear correlation between the way in which advice sequences are set up and the responses they elicit from clients. Where the counsellor delivered advice without eliciting a perceived problem from the patient, there were only three cases out of 32 cases recorded where the patient showed any sign of uptake. On the other hand, in the 18 cases where the advice was given in response to a request which was either volunteered or elicited from the patient, there are only four cases in which the patient did not show uptake.

It is clear that, contrary to the claims of some, there is no necessary contrast between the logic underlying qualitative and quantitative research. While much qualitative research is largely inductive and hypothesis-generating, this can be seen as reflecting a concern to avoid the imposition of *a priori* theoretical frameworks, which do not have a good fit with the empirical data to which they relate. However, as a body of theory develops, which is well grounded in the empirical world, there is no reason why the cumulative project should not be pursued through the subjection of such theory to new empirical test in the same way as Silverman *et al* (1992) have done. As Hammersley (1992g) concluded: "Which of these approaches is most appropriate should depend on our purposes, and the stage that our research has reached, not on paradigmatic commitments" (169).

3.2.1.3 Naturalism versus artificiality

A third divide which is sometimes postulated between qualitative and quantitative research is that the former investigates naturally occurring settings, while the latter is restricted to phenomena that are artificially created by the researcher. Once again, there is some truth in this characterisation of qualitative and quantitative research. Some qualitative researchers have indeed represented

[36] See section 3.2.2.5; [37] See appendix 1 for details of this study; [38] See appendix 1 for details of this study; [39] See appendix 1 for details of this study.

their own work as **naturalistic**, and implicitly or explicitly contrasted it with the **artificiality** of quantitative research. Indeed, the terms **naturalism** and **naturalistic** are sometimes used as a way of distinguishing qualitative from quantitative research (Lofland, 1967; Denzin, 1971; Schatzman and Strauss, 1973; Lincoln and Guba, 1985).[40]

Advocates of naturalistic research argue that the researcher should seek to remain faithful to the phenomena (s)he is studying (Hammersley and Atkinson, 1995). Phenomena should be studied in their natural state. This, they argue precludes the use of artificial research methods, such as experiments and structured interviews. The researcher is required to adopt an attitude of respect or appreciation for the people or setting under study. This concern to respect the world which is being studied is explicit in Blumer's writing: "The procedures employed in each part of the act of scientific inquiry should and must be assessed in terms of whether they respect the nature of the empirical world under study – whether what they signify or imply to be the nature of the empirical world is actually the case" (Blumer, 1969:27–8).

Similarly, Matza (1969) advocated **loyalty** and **fidelity** to the world being studied (5). Hammersley (1989) summarised this naturalistic position: "Central to naturalism is the desire to represent the world as it is, in all its complexity and changeability, and to avoid imposing artificial structures" (157).

This equation of qualitative research and naturalism has been extensively criticised, even among those who are sympathetic to qualitative research (Hammersley, 1992g; Silverman, 1989, 1993). Both Silverman (1989) and Hammersley (1992g) have argued that the distinction between artificial and natural settings is spurious:

> What happens in a school class or a court of law, for example, is no more natural (or artificial) than what goes on in a social psychological laboratory. To treat classrooms or courtrooms as natural and experiments as artificial is to forget that social research is itself part of the social world, something that should never be forgotten. (Hammersley, 1992g:164.)

Hammersley suggested that those who argue that qualitative research is less artificial than quantitative research are in fact confusing the issue. The difference between much qualitative research and quantitative research lies not in the artificial/natural divide, but in the extent to which the researcher seeks to **structure** the research situation. Whereas both experimenters and survey researchers seek to create special settings in which to carry out their research, and then play a dominant role in such settings, qualitative researchers are more likely to investigate naturally occurring settings and to seek to adopt a less intrusive and more marginal role in that setting (Hammersley, for the DE304 Course Team, 1979b).

As Hammersley (1992g) suggested, the argument in favour of studying naturally occurring settings, with minimum intrusion from the researcher, is generally constructed in terms of reducing the reactivity associated with quantitative methods of data collection (Duffy, 1987). Emerson (1983) cited Howard Becker's defence of field research in terms of their reduction of reactive effects. By studying people in their natural habitats, Becker argued, field researchers reduce the risk that their findings are artefacts of the experimental or interview situation. The people being studied:

> ... are enmeshed in social relationships important to them, at work, in community life, wherever. The events they participate matter to them. The opinions and actions of the people they interact with must be taken into account, because they affect those events. All the constraints that affect them in their ordinary lives continue to operate while the observer observes. (Becker, 1970:46 cited in Emerson, 1983:101.)

While we might accept that experiments and surveys are indeed highly prone to reactive effects, it does not follow that qualitative research is immune to reactivity. However passive and unobtrusive the researchers may be in the setting under study, their mere presence alters the setting in ways which may be significant (Hammersley, 1992g). Researchers, whether quantitative or qualitative, are inevitably part of the social world which they study (Hammersley and Atkinson, 1995). Indeed many qualitative researchers no longer aspire to such passive roles in the settings they are studying (Gussow, 1964; Cassell and Wax, 1980; Emerson, 1981; Hammersley and Atkinson, 1995). Rather they seek to treat their

[40] This use of the term 'naturalism' must be distinguished from the philosophical use of the same term. In the latter case, naturalism refers to a commitment to the unity of scientific method, or the belief that the same method or logic of explanation can be used in the social and natural sciences. Ironically 'naturalistic inquiry' (in the sense used by Lincoln and Guba, 1985 and others) is philosophically anti-naturalist.

impact upon the settings they study as a resource rather than a liability in their research:

> The investigator is conceptualized as part of the reality being studied. Here, it is taken for granted that the observer alters that which is observed; but these alterations are the subject of study ... in the first [view], the alterations, resulting from the effect of the observer upon that which is observed, are interference; in the second, they are data. (Cassell and Wax, 1980: 261, cited in Emerson, 1981:365)

Emerson (1981) observed that, far from adopting the passive role of earlier field work, some more recent qualitative researchers have advocated a more fully participatory role in field research.[41]

Once again, this suggests that the posited dichotomy between reactive quantitative and non-reactive qualitative research is misleading. Clearly, unacknowledged reactive effects in both qualitative and quantitative research may seriously undermine the ecological validity of conclusions drawn from the data. The critical issue is not whether or not such reactive effects exist, but whether they are taken into account in the analysis of the data.[42]

3.2.2 Qualitative versus quantitative research practice?

Having considered some of the philosophical issues underlying the so-called qualitative–quantitative debate, we now turn to consider some aspects of research practice which are presented by some as definitive of qualitative research. In doing so, we shall structure our discussion around the six features which Bryman (1988) identified as characteristic of qualitative research:

- commitment to viewing events, actions, norms, values etc. from the perspective of those being studied
- emphasis upon the description of the setting being investigated
- emphasis upon context and holism
- emphasis on process
- flexibility of research design
- reluctance to impose *a priori* theoretical frameworks at the outset.

Bryman's list should be seen as merely an organisational device. We do not intend that our use of this list should be understood as an endorsement of the view that these six features are either necessarily or actually definitive of qualitative research. Rather, just as we have argued that there is no necessary connection between any of the positions outlined above and qualitative or quantitative method, so we will show that many of the items on Bryman's list are contestable and are, in fact, contested (Silverman, 1993).

3.2.2.1 Commitment to viewing events, actions, norms, values etc. from the perspective of those being studied[43]

Bryman (1988) identified this as being the most fundamental characteristic of qualitative research. He elaborated: "Such an approach clearly involves a preparedness to empathize, (though not necessarily to sympathize) with those being studied, but it also entails a capacity to penetrate the frames of meaning within which they operate" (61).

Other authors who have presented such understanding as central to qualitative research include Wiseman (1970), Lofland (1971), Gould and co-workers (1974), Patton (1980), Duffy (1985), Marshall (1985), Finch (1986), Duffy (1987), Merriam (1988), Hammersley (1990), Habermann-Little (1991), Henwood and Pidgeon (1993), Oiler Boyd (1993a) and Lindlof (1995). Finch (1986) argued that it is this emphasis upon making settings and human actions comprehensible that makes qualitative research so relevant to the concerns of policy makers. For example, in the health field, much of the recent work on lay health concepts has served to show how seemingly irrational behaviour can be understood as highly rational, given the assumptions that patients are making. Similarly, Jensen (1989) has pointed to the practical implications of recognising that what people believe to be true may be more important than any so-called objective reality, given that people can be expected to act on the basis of what they believe to be the case. These authors reflect Thomas's often cited maxim that "If men define situations as real, they are real in their consequences" (Thomas and Thomas, 1927:572).[41]

Hammersley and Atkinson (1995) suggested that this commitment to understanding members' perspectives implies that, in practice, the qualitative researcher seeks to employ methods that facilitate access to members' meanings, rather than obfuscating them. Similarly, Henwood and Pidgeon (1993) argued that qualitative researchers seek to

[41] See section 4.2.2 for a discussion of the range of roles open to the participant observer; [42] See chapter 5;
[43] See section 2.2 and 2.3 for an historical perspective on the risks associated with ignoring the role of purposive action in explanations of human conduct and social events; [44] See section 2.4.3.

avoid the tendency that quantitative research has to fix meanings without reference to the meanings employed by participants in context.

Atkinson (1979) linked this emphasis upon understanding to the assumptions which qualitative researchers make about the way in which the social and natural world differ from one another:

> The social world differs from the natural world because it is **essentially** a world of interpretations and meanings ... People differ from natural objects in their ability to interpret their own actions and those of others, to act on their understandings and to endow their lives and actions with meaning. The social world of a particular culture is, therefore, **socially constructed**; it is the active accomplishment of the members of that culture. For this reason the language of ethnography refers to **actors** and **actions**, rather than, say, subjects and behaviour, and the question is always, 'How is it done?'; 'What cultural resources, stocks of knowledge, routines and strategies do the actors bring to bear?'; 'How do the actors collectively negotiate and achieve social order, understanding and working relationships?' (46 original emphasis).

Agar (1980) also emphasised the importance, within qualitative research, of learning to see the world through the eyes of the participants:

> The social research style that emphasises encountering alien worlds and making sense of them is called **ethnography** or 'folk description'. Ethnographers set out to show how social action in the world makes sense from the point of view of another. (12 original emphasis)

Emerson (1983) made the link between this concern with meaning in qualitative research and the distinction which Schutz (1962) made between the subject matter of natural and social science:

> The world of nature, as explored by the natural scientist, does not 'mean' anything to molecules, atoms, and electrons. But the observational field of the social scientist – social reality – has a specific meaning and relevance structure for the human beings living, acting and thinking within it. By a series of common-sense constructs they have pre-selected and pre-interpreted this world which they experience as the reality of their daily lives. It is these thought objects of theirs which determine their behaviour by motivating it. (Schutz, 1962:58–9 cited in Emerson, 1983:14.)

Fetterman (1989) suggested that the commitment to understanding members' perspectives relates to what anthropologists refer to as the **emic** perspective. Fetterman defined this perspective:

> The emic perspective – the insider's or native's perspective of reality – is at the heart of most ethnographic research. The insider's perspective of reality is instrumental to understanding and accurately describing situations and behaviours. Native perceptions may not conform to an 'objective' reality, but they help the field worker understand why members of a social group do what they do. In contrast to *a priori* assumptions about how systems work from a simple, linear, logical perspective – which[45] might be completely off target – ethnography typically takes a phenomenologically oriented research approach. (30)

Emerson (1983) criticised the implication that emic approaches involve the description of a culture in its own terms. He argued that the constructs underlying emic accounts cannot be seen as literally members' constructs. Rather they are necessarily "second order renderings of those constructs produced in one fashion or another by the ethnographer" (Emerson, 1983:24). Here Emerson was drawing on Schutz's distinction between first and second order constructs:

> The constructs used by the social scientist are, so to speak, constructs of the second degree, namely constructs of the constructs made by the actors on the social scene, whose behaviour the scientist observes and tries to explain in accordance with the procedural rules of his science. (Schutz, 1962:5–6.)

The commitment to understanding behaviour as **meaningful** is often associated with a preference for methods which all the researcher to get close to the people (s)he is studying. This may involve spending extended periods in the setting under study, as in participant observation.[46] Indeed, Denzin (1970) has presented this characteristic as definitive of participant observation: "Participant observation is a commitment to adopt the perspective of those studied by sharing in their day to day experiences" (185).

Foster (1993) identified the participant observer's capacity for seeing the world from the perspective of those studied as a major advantage of this method:

> The other main advantage [other than reducing reactivity] of this role is that the researcher is better able to see the social world from the point of view of his or her subjects. He or she has to learn the culture in order to operate as far as possible as an insider, and gains access to information not available to outsiders. In this way, the researcher is more likely to appreciate

[45] See section 4.3 for a discussion of standardised and non-standardised interviewing; [46] See section 4.2.

> and understand the subjects' perspectives and the meanings which underpin their interaction. In short, the researcher can put him- or herself in their shoes. (55)

In-depth interviewing techniques, which encourage informants to express their understandings and perspectives in their own words and reflecting their own priorities, are generally preferred to the practice of asking informants to respond to survey designers' assumptions about what is or is not important in the context of the research.

Bryman (1988) compared this approach within qualitative research with the orientation which is typical of quantitative research:

> The quantitative researcher adopts the posture of an outsider looking in on the social world. He or she applies a pre-ordained framework on the subjects being investigated and is involved as little as possible in that world. This posture is the analogue of the detached scientific observer ... Among qualitative researchers there is a strong urge to 'get close' to the subjects being investigated – to be an insider. For qualitative researchers it is only by getting close to their subjects and becoming an insider that they can view the world as a participant in that setting. (96)

Implicit in this orientation towards members' points of view within qualitative research, is a criticism of quantitative research: "Quantitative researchers seldom are able to capture the subject's perspective because they have to rely on more remote, inferential materials" (Denzin and Lincoln, 1994b).

The argument that qualitative research is committed to uncovering participants' meanings is seriously contested by many qualitative researchers. Such authors raise a number objections to the position discussed above where the objective of qualitative research is defined as the uncovering of participant meanings.

First, a number of authors (Emerson, 1981; Hammersley, 1992g; Silverman, 1993) argue that it is factually inaccurate to suggest that a focus on participant meanings is definitive of qualitative research. Hammersley (1992g) argued that, while much qualitative research does seek to understand the participant's point of view, this should not be seen as the end point of qualitative inquiry. Rather the aim is often to understand participants' behaviour. Thus understanding participants' perspectives is more appropriately understood as a means to that end, rather than as an end in itself. He added:

> It is very rare for qualitative research to restrict itself to documenting the native's point of view. And there are good reasons for not doing this; not the least of which is that the people studied can often do this for themselves! (165)

Similarly, Silverman (1993) argued that a characterisation of qualitative research as primarily concerned with uncovering meaning is inadequate because it does not do justice to the variety of qualitative research, some of which emphasises participant **practices** rather than participant **meanings**. Emerson (1981) again pointed out that much qualitative research, particularly that which is oriented towards ethnomethodology,[47] is less concerned with participants' perspectives than with their practices:

> Ethnomethodologically oriented field work is more concerned with actors' practices and practical concerns (Zimmerman 1970) than with their perspectives and cognitive categories. Actors attend to the often mundane 'tasks at hand', attempting to realize immediate 'practical purposes' that tend to be logically contradictory when viewed as parts of a single rational plan ... 'Perspective is both too cognitive and too subjective in its connotations: it leads the analyst to slip 'in and out of points of view', treating 'meanings of objects as more or less freely conjured' (Bittner, 1973:121–2). (Emerson, 1981:358.)

Many qualitative researchers are more concerned with studying the **function** rather than the meaning of practices (Silverman, 1993). Silverman and co-workers (1992)[48] offered an example of such a functional analysis, using their own study of HIV/AIDS counselling, which was discussed in section 3.2.1.2, as an example. This study was based upon an analysis of audio-tapes of counselling interactions in ten medical centres. They found that advice-giving to patients was typically truncated and non-personalised – features which may be expected to be related to low patient uptake. Silverman and co-workers considered why it is that the counsellors appear to prefer to package their advice to patients in a way which is likely to minimise patient uptake. Rather than criticising the professionals involved, Silverman and co-workers considered the functions which such demonstrably ineffective practices may serve within the context in which they are performed. They showed how such truncated and non-personalised advice packages can be effective in handling the

[47] See section 2.4.4 for a discussion of ethnomethodology; [48] See appendix 1 for details of this study.

delicate issues which arise in relation to sexual practices. The advice was presented in such a generalised way that it did not appear to single out the particular patient's sexual activities for discussion. Similarly, since they did not involve direct questioning about the individual patients' sexual activities or history, such practices meant that patients were not required to discuss their own behaviour. Again, these practices avoided the embarrassment associated with telling comparative strangers what they should be doing in their intimate lives. Finally, these practices minimised the likelihood of any overt conflict within the consultation. Silverman and co-workers concluded that, at least at one level, these practices were functional. Among other things, they might help the counsellor to limit the time which such counselling sessions require. On the other hand, they also involved serious losses, insofar as they were much less likely to be effective in persuading patients to reconsider their own sexual behaviours. It is only by understanding the positive functions which particular practices serve that the researcher is able to propose modifications of current practice which take account of such positive functions. As the authors observed, simply recommending ways in which counsellors could modify their behaviour to increase patient uptake, without supporting organisational change, could be expected to have minimal, or worse still, harmful, impact.

A second objection to defining qualitative research as preoccupied with uncovering participants' meanings is concerned with the **feasibility** of doing so. Emerson (1981) asked how we are to access participants' meanings. If researchers rely upon participants to explain their actions and interactions they immediately encounter two problems. First, there is the possibility that social actors have knowledge and experience which are relevant to the matter in hand but which they are unable to articulate. In such cases, as Emerson pointed out, their attempts to share their understandings may serve to distort and over-simplify. Second, and even more problematically, asking the participant to explain is likely to produce an 'account' (Scott and Lyman, 1968; Scott and Lyman, 1970; Lyman and Scott, 1989). The status of such accounts is more fully discussed in section 4.3. Here we simply note that, as Gould and co-workers (1974) pointed out, such accounts are devised to "make the system meaningful to an outsider" (xxiv) and bear an indeterminate relationship to the participant's decision to act in one way rather than another.

A third objection to the argument that the aim of qualitative research is to report participants' meanings relates to the status which is to be given to such reports. As both Hammersley (1992g) and Bryman (1988) pointed out, researchers do not simply **reproduce** participants' perspectives. Bryman (1988) observed that qualitative researchers are inevitably selective in the aspects of members' perspectives on which they focus. In addition, they "rarely adopt a stance of being 'sponges' whereby they simply absorb subjects' interpretations" (73). Rather they bring to bear their own theoretical perspectives which, at the very least, filter their observations. Here Bryman was drawing upon Bittner's observation that it is simply impossible for research to be undertaken in such a way that it is objective and devoid of presuppositions (Bittner, 1973).[49] As Emerson (1981) observed: "Inevitably field workers comprehend, interpret and code in memory ongoing social activities in ways that depend upon their presupposition, general cultural knowledge, prior experience in and particular knowledge of the setting" (358).

This raises, as a central issue of concern, "the feasibility of seeing through others' eyes if observers themselves are so heavily implicated in what is found" (Bryman, 1988:77).

In attempting to present participants' interpretations, researchers are inevitably involved in interpreting members' interpretations (Bittner, 1973; Douglas, 1976; Emerson, 1981). The researcher does not simply produce a description of the participant's point of view. Rather (s)he produces a description of the participant's point of view from the point of view of the social scientist who is the observer (Bittner, 1973). Thus, as Hammersley (1992g) has suggested, even where the ostensible focus is restricted to studying participants' perspectives, these are not simply reproduced. Rather the qualitative researcher will typically " seek to analyse their structure and/or production in ways that are likely to be alien to the people studied" (165).

These objections to treating participants' meanings as the topic of research should not be taken to imply a behaviourist position which excludes any recognition of the meaningfulness of behaviour. The critics of exploration for meaning approaches

[49] See section 2.3.1 for a discussion of the historical antecedents of the insistence upon the theory-laden nature of observation.

to qualitative research generally accept that the object of social science is **socially meaningful behaviour** rather than human behaviour viewed as a physiological process (Emerson, 1983:23). As Wieder (1985) put it:

> Our everyday experiences present us not with mere bodies, but with embodied consciousness. That which is bodily given appresents[50] to us a subjectively shared, intersubjective surrounding world. (81)

Mason (1996) incorporated Silverman and Emerson's emphasis upon practice within her broad characterisation of qualitative research as interpretivist, while pointing to the range of versions of such interpretivism which are operative within qualitative research:

> [Qualitative research is] grounded in a philosophical position which is broadly 'interpretivist' in the sense that it is concerned with how the social world is interpreted, understood, experienced or produced. Whilst different versions of qualitative research might understand or approach these elements in different ways (for example, focussing upon social meanings, or interpretations, or practices, or discourses, or processes, or constructions) all will see at least some of these as meaningful elements in a complex – possibly multi-layered – social world. (4)

However, critics of the exploration for meaning approach to qualitative research argue that the problem of interpretation, of knowing other minds, is not as straightforward as some would have us believe. We are not presented with the meaning which an event or an action has for participants in any direct fashion. Rather, our only access to the meaningfulness of others' behaviour is indirect. Schutz (1962) made this point clearly: "If we disregard the phenomenon of telepathy ... knowledge of another's mind is possible only through the intermediary of events occurring on or produced by another's body" (313–4).

Reporting the behaviour of research participants necessarily involves giving that behaviour meaning or making sense of that behaviour. However, rather than treating the meaning of such behaviour as self-evident or relying upon participants' accounts to give access to what is really going on here, the researcher is required to present the evidence upon which his or her interpretations are based (Mehan, 1978; Bryman, 1988; Mason, 1996). Emerson (1981) attributed the increased interest in how researchers have arrived at their interpretations of the actions of others to the development of ethnomethodological critiques of the neo-Chicago School.[51] Ethnomethodologists have emphasised the importance of the field worker's explication of the "commonsense knowledge and interpretive procedures he/she used to understand others' behaviour" (358).

As Bryman (1988) observed, this is frequently not done in the reporting of qualitative research:

> There is a tendency towards an anecdotal approach to the use of 'data' in relation to conclusions or explanations in qualitative research. Brief conversations, snippets from unstructured interviews, or examples of a particular activity are used to provide evidence for a particular contention. ... Further, field notes or extended transcripts are rarely available; these would be very helpful in order to allow the reader to formulate his or her own hunches about the perspective of the people who have been studied and how adequately the ethnographer has interpreted people's behaviour in the light of the explication of their systems of meaning. (77)

Mehan (1978) presented what he considered good practice in this respect:

> For constitutive ethnographers, exhaustive data treatment is a necessary check against the tendency to seek only evidence that supports the researchers' orienting hypotheses or domain assumptions ... Constitutive studies therefore attempt an exhaustive analysis of behaviour in the flow of events. (36–7)

The theory-impregnated and selective character of description is further discussed in section 3.2.2.2. The importance of the presentation of the evidence upon which interpretations are based is discussed in section 5.3.4.

This review suggests that, contrary to the claims of a number of authors, any attempt to define qualitative research in terms of a commitment to viewing events, actions, norms and values from the

[50] '**Appresentation** is a technical term drawn from Husserl's writings on phenomenology. Wieder (1985) defines it in this way: "In Husserl's thought, appresentation refers to the coupling of two things in such a way that one experiential thing that is 'presented' or 'directly experienced' motivates or leads the experiencing subject's thoughts to further experiential positing of something else as present. Even though this something else is not experienced firsthand, it is experienced as being there alongside the strictly presented object, and, generally, it contributes sense to that which is strictly presented" (78); [51] See section 2.4.4.

perspective of those being studied does not do justice to the range of positions currently adopted by those carrying out qualitative research. While many qualitative researchers do associate themselves with this position, there is a significant body of qualitative research, in the field of health as elsewhere, which distances itself from such exploration of meaning approaches. The implications of this diversity within qualitative research are more fully explored in the case study on *Using qualitative research methodologies to study medical information systems,* in chapter 6.

3.2.2.2 Emphasis upon description of the setting being studied

A number of authors, as Bryman (1988) suggested, see description as the central aim of qualitative research (Sandelowski, 1986; Crabtree and Miller, 1991; Habermann-Little, 1991; Henwood and Pidgeon, 1993; LeCompte and Preissle, 1993a; Munhall, 1993a; Oiler Boyd, 1993a; Denzin and Lincoln, 1994b; Hammersley and Atkinson, 1995). For some, this has meant prioritising description over explanation. At times, this is linked to an assertion that explanatory research should not be attempted before a firm descriptive base has been established, on which explanatory studies can be built. For example, Crabtree and Miller (1991) argued that many areas of general practice still need fundamental descriptive research, and, in particular qualitative descriptive research. For others, the preference for description is associated with a re-definition of explanation in the social sciences. Here the argument is that description **is** explanatory. For example, LeCompte and Preissle (1993a) argued that ethnography is primarily descriptive, insofar as it is concerned with studying the interplay among naturally occurring empirical variables rather than with measuring the outcome of experimental manipulations. They expected that such ethnographic description would include the systematic identification of possible causal and consequential factors which are related to an event. However, as the authors acknowledged, it will not allow prediction in the same way as experimental research.

In such description there is an emphasis upon mundane detail, what Bryman (1988) described as "the apparently superficial trivia and minutiae of everyday life" (63). Such details are seen as important because they assist us in understanding what is going on in a particular context. As such there is a clear link between this emphasis upon description and the previous discussion of the importance of subjective meaning in qualitative research. Emerson (1981) defined this link: "To the extent that the initial (and sometimes primary, or sole) task of field work is seen as discovering the actors' perspectives or subjective meanings, field research takes on a strongly **descriptive** emphasis" (355).

Hammersley (1992b) recognised the value of careful description of social settings or groups:

> Nor is the value of description limited to what is, for us, the exotic. We often discover that there are features of event the most familiar settings of which we are unaware, recognition of which may subtly, even dramatically, change our understanding of those settings. Much recent ethnographic work in sociology has been concerned with 'making the familiar strange' in precisely this manner. (33)

Hammersley (1992b) identified three reasons for the emphasis on description in qualitative research:

1. The inductivist orientation of much qualitative research[52] emphasises the emergence of theory from description of the events under study. This means that description is likely to be viewed, at the very least, as a vital first stage in developing theory.
2. Much interest in ethnography relates to the description of events of which the reader is unlikely to have had any first-hand experience, and the vicarious access that the researcher offers to settings which might otherwise be inaccessible.
3. The emphasis in qualitative research upon the importance of context is associated with a concern to offer a full description to the reader.[53]

A number of authors have indicated that ethnographic description is more than mere description. It has variously been described as **thick description** (Geertz, 1973; Merriam, 1988), **rich description** (Denzin and Lincoln, 1994b), **analytic description** (McCall and Simmons, 1969; LeCompte and Preissle, 1993a), **in-depth description** (Marshall and Rossman, 1989) and **faithful description** (Sandelowski, 1986). Such authors were seeking to distinguish between the kinds of description which are characteristic of social science and those which are routinely engaged in by everyday participants in a situation.

[52] See section 3.2.1.2; [53] See section 3.2.2.3.

Hammersley (1992b) has examined the assumptions that underpin these attempts to identify the characteristics of such social scientific description. First, he examined the claim that such description is analytical or theoretical. He argued that there is substantial ambiguity in such claims. As he pointed out, theories are clearly different from descriptions. While descriptions deal with particulars, theories are about universals. Again, there is no such thing as an **atheoretical** description. As we discussed in section 3.2.2.1, all descriptions involve theoretical assumptions.[54] Hammersley argued that the concept of theoretical description is fundamentally flawed. He examined a number of interpretations of this concept and concluded that the most useful one is the idea that ethnographic descriptions are theoretical insofar as they involve the application of existing theories.

Hammersley (1992b) then proceeded to discuss the extent to which it is appropriate to see qualitative research as concerned with producing faithful descriptions. He associated this concern with the commitment of many qualitative researchers to naturalistic description.[55]

> The most common conception of the descriptive character of ethnographic accounts is that they map the morphology of some area of the social world. Indeed, this is often a key feature that advocates emphasise about ethnography in contrast to other approaches to social research. Practitioners of these other approaches are criticised for failing to investigate 'naturally occurring' phenomena in a sufficiently direct and detailed manning, failing to get beyond the 'veil' of their own commonsense assumptions (Blumer, 1969:39). (Hammersley, 1992b:23.)

Hammersley characterised such attempts at naturalistic description as the reproduction model of research, which was discussed in section 3.2.2.1:

> The aim being to investigate and describe the social realm as it really is, beyond all presumptions and prejudices. Culture, social systems or social worlds are assumed to be objectively existing phenomena present in the world and awaiting description ... Implicit in the reproduction model is the idea that there is one true description that the ethnographer's account seeks to approximate, albeit one that incorporates the multiple accounts of participants. (Hammersley, 1992b:23–4.)

Hammersley criticised this reproduction model of ethnographic research. He argued that empirical phenomena are descriptively inexhaustible and that it is possible to provide multiple true descriptions of any scene. What is included in a particular description reflects what is taken to be relevant:

> In framing descriptions, then, we cannot be concerned solely with truth: what is to be included in the description must also be determined by assumptions about what is relevant. These assumptions are partly based on theoretical ideas about the relationships among different types of phenomena, but ultimately on purposes and the values that ground these purposes. The idea that ethnographic accounts are simply descriptions of reality 'as it is' is just as misleading as the notion that historical accounts simply represent past events. (25)

Hammersley drew the implication that, since descriptions are always, and necessarily, from a particular, value-based point of view, it is the responsibility of the analyst to make the value and factual assumptions which underpin such descriptions explicit and to justify them, where it is necessary to do so.[56]

All of this suggests that much qualitative research is indeed concerned with description. Such description has particular value in challenging or undermining taken-for-granted or routine assumptions about the nature of the setting or group under study. Such description may helpfully highlight features of familiar social settings which are not otherwise recognised by participants in that setting. These can challenge "preconceptions we bring to our research and which so easily get built into the accounts we produce" (Hammersley, 1992a:33). However, as Hammersley pointed out, this stress on the value of description must be tempered with a recognition that such description can never be treated as a direct reproduction of reality. There is no such thing as pure description. Rather descriptions are always and necessarily imbued with theoretical assumptions which need to be made explicit in the account which the analyst gives of the setting or group studied.

3.2.2.3 Emphasis on context and holism

Bryman (1988) argued that contextualism and holism were almost inseparable from one another. He defined contextualism as a preference for "understanding events, behaviour etc. in their context" (64). Holism, on the other hand "entails an undertaking to examine social entities – schools, tribes, firms, slums, delinquent groups,

[54] See section 2.3.1 for an historical perspective on the possibility of atheoretical description; [55] See section 3.2.1.3 for a more detailed discussion of naturalism in qualitative research; [56] See section 5.3.5.

communities or whatever – as wholes to be explicated and understood in their entirety" (64).

This emphasis upon context and holism in qualitative research leads to a style of research in which, rather than attempting to isolate and manipulate variables, as in experimental and, to a more limited extent, survey research, the researcher seeks to study the phenomena of interest within the wider context in which they occur. Participant observation, qualitative interviewing and the analysis of documents in relation to the circumstances of their production,[57] are seen as having significant advantages in this respect.

As Bryman (1988) observed this strong emphasis upon the importance of understanding the phenomenon under study in the context of the culture, sub-culture, organisation or setting of which it forms a part is central to much qualitative research (Mishler, 1979; Aamodt, 1983; Duffy, 1985; Lincoln and Guba, 1985; Silverman, 1985; Merriam, 1988; Jensen, 1989; Stange and Zyzanski, 1989; Hammersley, 1990; Habermann-Little, 1991; Eakin and Maclean, 1992; Hammersley, 1992b; Hammersley, 1992c; Henwood and Pidgeon, 1993; LeCompte and Preissle, 1993a; Oiler Boyd, 1993b; Guba and Lincoln, 1994; Miles and Huberman, 1994; Baum, 1995; Clarke, 1995; Dootson, 1995; Keen and Packwood, 1995; Stake, 1995; Miller, 1997). Atkinson (1979) traced the emphasis on holism in qualitative research to historical roots in the development of social anthropology in the early twentieth century. He identified the emphasis upon context and holism as one of the features which set the social anthropology of Malinowski and others apart from nineteenth century theorists, such as Frazer, who focused upon isolated features of culture, rather than seeing them as part of their social context.[58] The insistence upon relating social phenomena to their context is one of the features of qualitative research that is seen to be in direct contrast to most quantitative research.

Much quantitative research is concerned to isolate causal relationships that are believed to under-pin social reality. In attempting to identify causal relationships, and ultimately universal laws, which are, by definition, context-free, quantitative analysts seek to isolate dependent and independent variables. Such analysts recognise the difficulties that are posed by the complexity of the social world and the ever-present risk of spurious correlations. Stake (1995) argued that quantitative researchers do everything in their power to nullify context, try to eliminate the merely situational and treat uniqueness as error. This is associated with a preference for research methods that incorporate either physical or statistical controls and the elevation of the randomised controlled trial (RCT) to the peak of the methodological hierarchy. It is only by such control that it is possible to isolate cause and effect and to develop theories that have predictive power. Mishler (1979) has argued that experimental design depends upon removing people from their usual contexts so as to be able to hold constant potentially confounding variables. The methods of quantitative social science are directed towards isolating variables from their personal and social contexts.

Qualitative researchers engage with the complexity of the social world somewhat differently (Waitzkin, 1990). Rather than trying to control such complexity, they argue that it should be placed at the centre of research. They argue that the reductionist procedures of quantitative research are inadequate for unpacking the complex mix of social, economic, political and environmental factors, which are at the heart of human action (Baum, 1995). Such procedures "ride roughshod over the complexity of the social world" (Hammersley, 1992c:32). They fail to take account of the fact that social variables are intrinsically more difficult to isolate and test than those in the natural sciences (Silverman, 1985). Mishler (1979) coined the term **context-stripping** to describe this approach. Clarke (1995) argued that such context-stripping approaches fail to recognise that behaviour can only be fully understood in context. Duffy (1985) argued that the emphasis upon the control of extraneous variables in experiments has the effect of ruling out the kind of interactive effects which are typical of non-experimental situations. Cochrane (1972), himself, made this same point:

> Between the scientific measurements based on RCTs and the benefit measurements ... in the community, there is a gulf which has been much underestimated ... Different strategies of management may be needed to reach the levels of effectiveness comparable to those reached in the RCTs. There is in addition the vast problem of the optimum use of personnel and materials in achieving those results. (Cochrane, 1972:2.)

Guba and Lincoln (1994) summed up the objections to reliance upon such context-stripping approaches:

[57] See sections 4.2., 4.3 and 4.4; [58] See section 2.4.1.

> Precise quantitative approaches that focus on selected sub-sets of variables necessarily 'strip' from consideration, through appropriate controls or randomization, other variables that exist in the context that might, if allowed to exert their effects, greatly alter findings. Further, such exclusionary designs, while increasing the theoretical rigour of a study, detract from its **relevance**, that is, its applicability or generalizability, because their outcomes can be properly applied only in other similarly truncated or contextually stripped situations (another laboratory, for example). (106)

Some authors appear to object to context-stripping in social science because of the reductionist model of the individual which it assumes:

> Individuals are not reducible and measurable objects that exist independently of their historical, cultural and social contexts. To treat them as such, as quantitative research does, is to reduce them to machine-like figures that are only a sum of their parts. (Duffy, 1987:130.)

The emphasis upon context within qualitative research has been seen as particularly relevant to issues related to HTA. For example, Keen and Packwood (1995) argued that qualitative research is particularly appropriate in evaluation studies where the general context of an intervention or programme could be expected to be important in influencing the outcome of a programme.[59] Eakin and Maclean (1992) argued that it is this emphasis on context and holism that makes qualitative research particularly attractive to health promotion researchers who are disillusioned with the inability of conventional methods to take account of the complexity of behaviour in natural settings.

3.2.2.4 Emphasis on process

The fourth characteristic of qualitative research identified by Bryman (1988) is an emphasis upon the processual and dynamic nature of social life. This characteristic is also emphasised by a number of other authors (Delamont and Hamilton, 1976; Atkinson, 1979; Bogdan and Biklen, 1982; Finch, 1986; Duffy, 1987; Merriam, 1988; Hammersley, 1990; Strauss and Corbin, 1990; Dingwall, 1992; Steckler *et al,* 1992; Munhall, 1993b). Quantitative research is criticised for treating social phenomena as more clearly defined and static than they really are. Bryman (1988) linked this emphasis on process to a view of social life, which sees it as involving interlocking series of events that are in a state of flux and change. This is associated with a preference for longitudinal designs that are capable of capturing the processual aspects of a phenomenon. In particular, participant observation studies have a capacity for studying the way in which such phenomena change over time.

It is this emphasis on process which makes qualitative research particularly appropriate for studying the implementation of a programme (Finch, 1986). As such, it is relevant to those charged with developing or implementing policy. Finch argued that participant observation,[60] in particular, is well suited to studying process because it is not forced to rely upon retrospective accounts and because it is flexible enough to respond to unexpected developments. The emphasis on process marks qualitative studies out from conventional input–output studies. The critique of what has been referred to as the **black box** model of input–output research was developed particularly in the field of education in the 1970s. It can be seen as having equal relevance in the field of health care (Dingwall, 1992).

Atkinson (1979) described the emergence of this critique within education:

> Until the 1960's sociological research in the field of education all too often treated the school as a 'black box': researchers were generally content to measure the 'input (e.g. social class, family background and individual ability) and the 'output' (e.g. attainment and occupation), whilst the process of schooling remained largely unexplored. The Plowden research represents this kind of approach ... Bernstein and Davies (1969) ... noted that Plowden makes little attempt at detailed description of schools and pointed to its 'trivial' discussion of differences in schools, justified by the claim that 'what goes on in primary schools cannot greatly differ from one school to another, since there is a limited range of material within the capacity of primary school children'. Since the 1960's there has been increasing (though still fairly limited) research into school life and the process of schooling. Since then sociologists of education have begun to study the process of schooling and, in particular, face-to-face interaction in classrooms. (Atkinson, 1979:48.)

Within healthcare settings, qualitative research can be seen as representing a similar challenge to the input–output model. Steckler and co-workers (1992) argued that in the field of health promotion, there is a need for qualitative studies which provide valid process data, to complement reliable outcome data from quantitative studies. Similarly, Dingwall (1992) argued that, while recent health services evaluations in the UK have focussed

[59] See chapter 7; [60] See section 4.2.

primarily on outcome measures, such approaches are ill-suited to providing at least some of the information that policy makers and practitioners need. Such outcome studies, he argued, lack explanatory power. They can establish a link between A and B, but are unable to explain the process by which A was transformed into B. He suggested, therefore, that outcome studies need to be supplemented with process studies which are capable of capturing the dynamic aspects of an organisation. Such process studies must be carried out from within the organisation under study, using, for example, the methods of participant observation.[61]

Silverman (1989) pointed out that statistical analysis may conceal as well as reveal social processes. As Cronbach (1975) argued, such statistical analyses may provide correlations linking input variables to output variables. However, they are less equipped to help us understand the potentially important cases in which such generalisations do not hold. Qualitative, observational methods have particular strengths in terms of uncovering the social processes that intervene between input and output variables and identifying the reasons why the general relationship does not hold in particular cases.

3.2.2.5 Flexibility and lack of structure: reluctance to impose a priori *theoretical frameworks at outset*

Bryman (1988) identified qualitative research with a preference for a relatively open and unstructured research strategy. He associated this with a commitment to viewing social phenomena through the eyes of participants[62] and a consequent wariness about "imposing prior and possibly inappropriate frames of reference on the people they study" (66). There are close links between such flexibility in research design and the reluctance of many qualitative researchers to impose pre-formulated theoretical frameworks and concepts in advance of the study itself. As discussed in section 3.2.1.2, a number of authors (most notably Glaser and Strauss, 1967) have argued that theories and concepts should emerge from the setting under study, rather than being imposed from outside.

Qualitative research is sometimes contrasted in this respect with quantitative approaches, which are accused of imposing the researcher's prior assumptions on those being studied, thus reducing the opportunity for discovering evidence that is discrepant with those assumptions (Hammersley, 1992b). This has been seen as a particular problem in health-related research insofar as researchers may be closely allied to doctors and other health professionals whose assumptions and understandings may be very different from those of patients or clients.

This insistence that the social world must be **discovered** is associated with a preference for research methods which involve first hand observation and participation. Relying on what people say, without observing what they do (as in survey research) is seen as particularly likely to distort social reality. Qualitative researchers argue that their methods allow them to discover social processes and social meanings through direct engagement with and immersion in the concrete reality of social life (Hammersley, 1992b).

These twin emphases on flexibility and discovery are seen as central to qualitative research by a number of authors (Aamodt, 1983; Charmaz, 1983a; Lincoln and Guba, 1985; Marshall, 1985; Duffy, 1987; Imle and Atwood, 1988; Stange and Zyzanski, 1989; Hammersley, 1990; Munhall, 1993b; Guba and Lincoln, 1994; Britten, 1995). Recently the concern not to impose inappropriate frames of reference has characterised much feminist qualitative research.[63] Writers such as DeVault (1990) and Griffen (1986) have argued that flexibility in research design is important if the researchers is to avoid imposing distorting frameworks, based upon male assumptions, upon the experience of women.

These authors do not necessarily assume that qualitative researchers will embark upon their research *tabula rasa.* However, any theories or concepts which they bring to a study will be held lightly and be subject to reformulation or rejection as the research progresses. Merriam (1988) summed up this commitment: "Occasionally one may have working hypotheses at the outset of a study but these expectations are subject to reformulation as the study progresses" (13).

Guba and Lincoln (1985) used the term **emergent design** to describe the process of qualitative research. By this they meant that, rather than being fixed at the outset, the design of qualitative research emerges as the study progresses, in response to the researcher's early observations. Atkinson (1979) argued that, in qualitative research:

[61] See section 4.2; [62] See section 3.2.2.1; [63] See section 3.1.3.

> Research design refers to a multitude of decisions that have to be taken **over the whole course of the fieldwork.** In each case the strategy that is adopted depends to a great extent on the nature of the social situation chosen for study. (45 original emphasis.)

This emergent design is closely associated with the notion of **progressive focussing**. Atkinson (1979) described how this works:

> The emphasis upon discovery requires research strategies with a wide focus, collecting any data which are possibly relevant. Ethnographers try to avoid sharpening their problems into specific research hypotheses until considerable exploratory investigation has occurred (a process termed progressive focussing). (53)

It is also associated with the practice of theoretical sampling.[64]

Strauss and Corbin (1990) stressed the importance of discovery when they summarised some of the reasons why a researcher might opt to use qualitative methods:

> Qualitative methods can be used to uncover and understand what lies behind any phenomenon about which little is yet known. It can be used to gain novel and fresh slants on things about which quite a bit is already known. Also qualitative methods can give the intricate details of phenomena that are difficult to convey with quantitative methods. (19)

Marshall (1985) presented the flexibility of qualitative research as one of its strengths. She argued that, in qualitative research, the researcher does not assume that (s)he knows, at the outset, the exact nature of the research question. Rather, qualitative research offers the opportunity of discovery. In fact, researchers vary considerably in the extent to which they formulate the precise nature of the research question in advance of the study. Most seek to avoid what Silverman (1993) referred to as "premature definition of variables".

However, perhaps partially at least, at the behest of funding bodies, qualitative researchers are often ready, albeit in fairly general terms, to define their research question in advance of data collection. Thus Bloor (1976)[65] identified his interest as establishing: "whether or not geographical differences in the incidence of adenotonsillectomy among children could be attributed to ... differences between ENT specialists in different geographical areas in their routine assessments" (545).

Similarly, Gantley and co-workers (1993)[66] defined the objective of their study of infant care practices among an ethnic minority population, which was discussed in section 3.2.1.2, relatively precisely: "To investigate the infant care practices of a small ethnic minority population within Britain that might suggest possible factors contributing to the low incidence of the sudden infant death syndrome in the Asian population" (16).

Like Bloor, these authors related their focus to the findings of previous quantitative research. In this case, the authors' point of departure was the quantitative, epidemiological evidence of international and intranational variations in the incidence of sudden infant deaths.

While their research question was relatively precise, their research strategies remained flexible. They described their approach as **social anthropological** and spelled out the way in which this was translated into a flexible, discovery-based research strategy:

> Social anthropology addresses cultural ideologies by using a variety of methods including community observation and ethnographic interviewing. ... Ethnographic interviewing describes an open-ended, unstructured approach designed to encourage informants not only to describe infant care practices, but also to locate them within a broader ideology. The interviewer sets the framework for the interview – in this case mothers were asked to describe a 'day in the life' of their infant – but allows the informant to determine the pace and order of the conversation, to select topics considered important. (17)

As discussed in section 3.2.1.2, the emphasis on flexibility in research design is sometimes associated with the idea that qualitative research is best suited to theory generation, rather than theory testing (Glaser and Strauss, 1967; Aamodt, 1983; Kirk and Miller, 1986; Sandelowski, 1986; Merriam, 1988; Stange and Zyzanski, 1989; Beck, 1993; Henwood and Pidgeon, 1993; Munhall, 1993b), or the context of discovery rather than the context of justification (Reichenbach, 1938; Reichenbach, 1951). For example, Beck (1993) argued for a division of labour between qualitative and quantitative research, such that the latter concentrates upon identifying the characteristics of phenomena, while the latter engages in prediction and hypothesis testing. Hammersley and Atkinson (1995) identified the assumption that the generation of theory belongs to the context of discovery, and hence, in Reichenbach's terms at least, lies outside

[64] See section 4.1; [65] See appendix 1 for details of this study; [66] See appendix 1 for details of this study.

the realm of science, as a distinctive feature of positivism.[67] Kirk and Miller (1986) argued that one of the weaknesses of most quantitative research lies in its lack of concern with discovery. They argued that the preoccupation with hypothesis testing in quantitative research has led to a lack of attention to the equally important task of generating hypotheses for testing. Even those who emphasise the priority of measurement may see a role for qualitative research in providing the concepts which may subsequently be translated into a measurement scale (Imle and Atwood, 1988).

Conversely, however, a number of authors, sympathetic to qualitative research, have challenged the view that the proper field of qualitative research is hypothesis generation rather than hypothesis testing. Their position was discussed section 3.2.1.2. They argue that, in qualitative research, the preoccupation with hypothesis generation has had the unfortunate effect that little effort has been directed towards building cumulative bodies of knowledge on the basis of qualitative research (Hammersley *et al,* 1985; Silverman, 1989). As Silverman (1993) argued, while qualitative research may, in the past, have been characterised by a lack of concern with cumulative knowledge, increasingly, qualitative research does begin with prior hypotheses/definitions (e.g. Silverman, 1984). The importance of cumulative, qualitative research is further discussed in chapter 5.

3.3 Summary and conclusion

In this section of the report we have reviewed the literature that deals with the relationship between qualitative and quantitative research, with particular reference to HTA. In section 3.1, we compared and contrasted two positions on this relationship, which can be discerned in the literature. On the one hand, there are those who argue that qualitative and quantitative research are based on two opposing and incommensurable paradigms. Their opposition is seen as lying in a fundamental philosophical between the *a priori* assumptions of each paradigm. On the other hand, there are those who argue that while both qualitative and quantitative research face significant problems in studying the social world, the choice between qualitative and quantitative methods for investigating this world should be made on instrumental rather than philosophical grounds.

In section 3.2, we considered the grounds on which advocates of the two-paradigms approach argue their position. We reviewed the argument that qualitative research is based upon an idealist, constructivist ontology, whereas quantitative research is based upon a realist ontology (section 3.2.1). We concluded that neither naive realism nor naive idealism do justice to the complexities which confront those who engage in social research. We concluded that the adoption of the radical constructionist position advocated by many of those who argue for the two-paradigms approach raises significant problems in relation to justifying the funding of such research within policy-oriented fields such as HTA.

We then considered the argument that the logic underlying qualitative and quantitative research is fundamentally different, with qualitative research adopting an inductive logic in comparison with the deductive procedures of quantitative research (section 3.2.2). We concluded that this categorisation of qualitative and quantitative research was not only inaccurate (in terms of the practices of those who use both qualitative and quantitative methods), but also both unnecessary and unworkable in practice. We then moved on to consider the argument that the distinction between qualitative and quantitative methods is best understood in terms of the commitment of the former to naturalism and the latter to artificiality (section 3.2.3). We concluded that, once again, this proposed dichotomy was spurious and runs the risk of distracting our attention from the reactive effects, which are an inevitable part of all social research, whether qualitative and quantitative.

Having concluded that there were no grounds for accepting that qualitative and quantitative research are either necessarily or indeed actually characterised by dichotomous philosophical assumptions, we moved on, in section 3.2.2, to consider the claims that the practices of qualitative and quantitative research are fundamentally different. In section 3.2.2.1, we considered the claim that qualitative research is characterised by a focus on participants' meanings. While accepting that qualitative research is characteristically concerned with meaningful behaviour, and that for many qualitative researchers this concern is translated into studies which seek to elicit the meanings that participants ascribe to their actions and the events in which they are involved, we concluded that there are serious difficulties with the argument that the proper focus

[67] See section 2.3.2 for a discussion of positivism.

of qualitative research is participants' meanings. We observed that, in reality, many qualitative researchers, in the field of health as elsewhere, focus on participants' practices rather than participants' meanings. We also concluded that understanding the meanings that underpin behaviour is more problematic than the advocates of exploration for meaning approaches in qualitative research sometimes suggest. In particular, we identified problems in relation to the reliance upon participants' accounts of the meanings that underlie such meaningful behaviour.

In section 3.2.2.2, we considered the claim that qualitative and quantitative research can be distinguished in terms of the greater emphasis the former places upon description. While accepting that much qualitative research does indeed prioritise description and that such description can play a significant role in undermining the taken-for-granted assumptions that may distort social research, not least in the health field, we concluded that any claim to pure description, in either qualitative or quantitative research is highly problematic. As much quantitative research is also best characterised as descriptive rather than explanatory, the concern with description in qualitative research cannot be used to distinguish qualitative and quantitative methods.

In sections 3.2.2.3 and 3.2.2.4 we moved on to consider the argument that qualitative research is particularly well suited to studies that are concerned with understanding events and behaviour, holistically, in the context in which they occur in everyday life, and to those that seek to understand the process by which such events and behaviours come about. We concluded that these features of qualitative research can justifiably be seen as characteristic of that approach. In keeping with an instrumentalist approach to the difference between qualitative and quantitative methods, we identified contextualism, holism and process as three of the particular strengths of qualitative research, and as features that should be taken into account in selecting methods which are appropriate to the study of particular phenomena.

Finally, in section 3.2.2.5, we considered the claim that qualitative research differs from quantitative research in terms of its research design and the reluctance to impose *a priori* theoretical frameworks at the outset. We concluded that, in practice, the research design of qualitative research is likely to be much more flexible than would be considered good practice in most quantitative work, and that, in certain circumstances this may bring considerable advantages. This again reflects an instrumentalist position on the relationship between qualitative and quantitative research. We also concluded that qualitative researchers are typically more reluctant to impose prior theoretical frameworks upon their data than their quantitative colleagues, often adopting the language of hypothesis generation rather than hypothesis testing. However, we observed that, on the one hand, a concern with hypothesis generation is not the exclusive preserve of qualitative researchers, and, on the other, an increasing number of qualitative researchers are turning their attention to hypothesis testing and the cumulative development of theory.

In conclusion, we have considered various dimensions of the arguments put forward by those who argue that qualitative and quantitative research should be seen as two incommensurable paradigms, marked by dichotomies of philosophy and practice. We have concluded that such arguments are ill-founded and, indeed, are likely to obstruct the useful application of qualitative methods to the challenges of HTA. We suggest that it is more profitable to recognise the complementarity of qualitative and quantitative methods, acknowledging the particular strengths of the latter in terms of their capacity for studying socially meaningful behaviour, holistically, in context and with due attention to the dynamic processual aspects of social events and interactions. In addition, the emphasis on flexibility of design is particularly well suited to hypothesis generation and discovery, and the reluctance to impose (as opposed to incorporate) prior theory has significant potential for undermining the taken-for-granted assumptions that might otherwise limit the usefulness and distort the findings of research in the field of HTA.

Implications for commissioning and practice

- Decisions about whether qualitative or quantitative methods (or some combination of the two) are most appropriate to a particular research problem in HTA should be made upon instrumental rather than philosophical or ideological grounds. Researchers should establish which approach is likely to answer the question at hand in the most effective and efficient manner.
- There is no necessary connection between qualitative research and particular ontological or epistemological positions. This means that it is possible to combine qualitative and quantitative research to address a research question in HTA.

The decision to do so should be based upon the contribution that each can make rather than upon the researchers' prior commitments.

- It is impossible, in either qualitative or quantitative research, to establish truth claims with absolute certainty. Nevertheless, the goal of all research in HTA should be to establish knowledge about which we can be reasonably confident and to provide findings that are relevant to policy makers and practitioners.
- There are some problems in HTA, as in other fields, which cannot be fully resolved using quantitative methods alone, and there are some circumstances in which qualitative methods represent the technically superior option. Qualitative methods are particularly suited to answering 'How does this come to happen?' questions rather than 'How many?', 'How much?' or 'How often' questions.
- There are some situations in which quantitative data is inaccessible. For example, carrying out a sample survey or intervention study among people involved in stigmatised or illegal activities might well prose major practical difficulties. Nevertheless, information about the impact of health technology among such groups may be of considerable importance for policy makers and practitioners. In such cases, qualitative methods may be the only practical (and therefore the preferred) option.
- Qualitative research is particularly useful in providing a rigorous descriptive base upon which subsequent explanatory research can be based. Such description is important in providing policy makers and planners with an understanding of the context in which policies will be implemented. It is also useful to practitioners insofar as it can highlight the mundane or taken-for-granted aspects of familiar settings, which may remain unnoticed by those operating in such settings, but may influence outcomes significantly.
- Qualitative methods are useful in the exploratory stages of a research project where they can help to clarify the research question, aid conceptualisation and generate hypotheses for later research. Sound qualitative research of this kind is important in avoiding the arbitrary imposition of *a priori* frameworks, which do not have a good fit with the empirical data to which they relate. Such preliminary qualitative research in HTA can improve the quality of subsequent research by avoiding the premature operational definition of variables and so promoting the efficient use of funds invested in quantitative research by promoting its effective execution.
- Qualitative research has particular strengths in uncovering evidence that is discrepant with researchers' or practitioners' prior assumptions. The flexibility, which is characteristic of qualitative design, permits the researcher to identify significant but unanticipated factors. This is a particular strength in health-related research insofar as researchers are often closely allied to doctors and other health professionals whose assumptions and understandings may be very different from those of patients or clients.
- Qualitative methods may be used to interpret, qualify or illuminate the findings of quantitative research. In particular, they are helpful in examining the exceptions which do not fit probabilistic relationships between variables that are established through quantitative research, thus leading to the refinement of models. They can be used, in HTA, to help us to understand the potentially important cases where generalisations do not hold.
- It is inappropriate, however, to limit qualitative research to the tasks of hypothesis generation or explaining unanticipated results. Hypotheses generated in qualitative studies can be tested using either qualitative or quantitative methods. The methods used should be tailored to the hypothesis under study, rather than being driven by prior commitments.
- Qualitative research has an important contribution to make whenever the context in which a health technology is to be implemented can be expected to have an impact upon the outcome of that technology. Used alongside experimental methods, qualitative research can improve the generalisability of findings by providing a detailed description of the context in which the technology was applied. In particular, qualitative researchers' concern with the impact of the research process itself upon the findings obtained has advantages for the application of findings beyond the research setting.
- Qualitative researchers seek to understand the phenomenon under study in the context of the culture, sub-culture, organisation or setting of which it forms a part. This increases the relevance of findings of such research insofar as it takes account of the kind of interactive effects, which are typical of the applied rather than the experimental situation. This is particularly important in assessing the impact of new health technologies where the general context could be expected to influence outcome as it allows researchers to take account of the complexity of behaviour in naturally occurring settings.
- Health technologies are applied by people (be they doctors, nurses, technicians or patients) to other people (usually patients). One of the

distinctive features of a **human** action is that it is meaningful. People act on the basis of what they believe to be true rather than what may be **objectively** true. Qualitative research resists the tendency, characteristic of some quantitative research, to fix meanings. By examining participants' behaviour in context, the researcher is in a position to draw inferences about such meanings, which may then be tested empirically. In particular, qualitative research may be useful in illuminating the factors which sustain professional practices that have been shown to be ineffective, inappropriate or harmful in healthcare settings.

- However, claims that qualitative methods allow researchers to discover why people behave as they do, using methods which rely upon self-reports, should be treated with caution. The reasons that people give for particular behaviours bear an indeterminate relationship to their decision to act in particular ways.
- Qualitative methods are particularly useful in illuminating the findings of outcome studies in HTA. Qualitative research can provide the information that policy makers and practitioners need to supplement the findings of conventional outcome studies. Statistical analysis may conceal as well as reveal social processes. While outcome studies can establish a link between intervention and outcome, they are less able to explain the process by which the intervention was translated into that outcome. It is these dynamic aspects of health technology that qualitative research is best able to illuminate.

Chapter 4

The methods of qualitative research

In this section of the report, we turn to a consideration of the methods used by qualitative researchers. We shall discuss the use of particular methods (such as observation and interviewing), as well as considering topics that apply to all qualitative methods (such as selection, analysis and ethics). We shall review the literature on the principal methods used in qualitative research and the debates that surround them. There is no intention to offer a guide to carrying out qualitative research. Rather we aim to outline the range of positions taken by qualitative researchers in relation to each of the methods and topics discussed, and to review the arguments that are put forward in support of these positions, with particular relevance to research in HTA. Where the methods used by qualitative researchers differ from those conventionally adopted by quantitative researchers, we discuss these differences and present the responses which qualitative researchers have made to their critics. Where appropriate, we also review debates within qualitative research.

4.1 Selection

One of the concerns most frequently expressed in relation to qualitative research, is that it fails to meet the criterion of generalisability,[1] which is often held to be the hallmark of science (e.g. Smith, 1975:88; Kennedy, 1979:663). Within quantitative research traditions, generalisability has traditionally been pursued via sample to population inference, using probabilistic sampling methods. Much qualitative research is carried out in a single setting, or with a small sample of informants and thus fails to meet the assumptions of the sample statistics upon which such inferences can be based (Miles and Huberman, 1994). As Hammersley (1992f) pointed out, a central issue facing qualitative researchers is how they are to generalise from findings about particular situations, which may in themselves have little relevance to readers, to conclusions which have general relevance. The relative absence of probability sampling within qualitative research has given rise to concerns that such research is likely to be merely anecdotal, and of no practical use to policy makers or practitioners.

Responses from qualitative researchers to this kind of criticism have varied. As Schofield (1993) has observed, earlier generations of qualitative researchers tended to disregard generalisability as "unimportant, unachievable or both" (201). Some, such as Denzin (1983) have rejected generalisability as an appropriate goal for qualitative research:

> For the interpretivist every instance of social interaction, if thickly described (Geertz, 1973), represents a slice from the life world that is the proper subject matter for interpretive inquiry ... Every topic must be seen as carrying its own logic, sense of order, structure and meaning. (133–4)

Similarly, Stake (1995) argued that, in case studies, "Seldom are we primarily trying to generalize to other cases" (134).

Schofield traced qualitative researchers' disregard for generalisability, in some measure, to the strong historical links between such research and anthropology.[2] In cultural anthropology, individual exotic cultures were often studied for their own sake, as a way of demonstrating the rich diversity of traditional cultures, sometimes with the goal of describing them before they disappeared. In such research, she argued, generalisation had little place. Indeed, it was the distinctiveness of cultures rather than what they shared in common, which was of particular interest (see also LeCompte and Preissle, 1993a). Schofield noted that, in recent years, many qualitative researchers have begun to take much more seriously concerns about generalisability and to seek to address them through their work. She attributed this change of climate partly to shifts in the uses to which qualitative research has been put, over the last two decades. In particular, in the field of education, qualitative research has been applied to both evaluation research and to more basic educational research, often funded by external agencies. The same is, of

[1] This is more fully discussed in chapter 5; [2] See section 2.4.1 for a discussion of the links between qualitative research and anthropology.

course true, in HTA, as evidenced by the commissioning of this review. In both kinds of work, issues of generalisability assume central importance, since funding agencies are generally less interested in particular settings, than in what light the study of such settings can throw upon other, comparable settings. Such generalisation beyond the setting immediately raises issues about the methods which were used to select the setting(s) studied. At the same time, the rapprochement between qualitative and quantitative research, and the rise in multi-method research has meant that qualitative researchers have been increasingly exposed to debates within quantitative research about generalisability and this has led to a reconsideration of these issues by many.

While some, including Hammersley (1992f), have suggested that occasionally a single case may have intrinsic interest in its own right, such that the representativeness of the case is of little concern, others, including Dingwall (1992), have suggested that: "The one-off case study, conceived and executed in magnificent isolation, has no place in modern social science and little more than anecdotal value to a policy maker trying to understand how an organisation works"(171).

4.1.1 Approaches to sampling in qualitative research

Most writers now accept that achieving findings that are of relevance beyond the settings from which they were initially derived is an important goal for both qualitative and quantitative research.[3] However, where researchers seek to apply the findings of their research, beyond the immediate context studied, the issue of case or sample selection becomes crucial. A number of different approaches to selection have been discussed within the literature on qualitative methods and, for ease of presentation, we have chosen to discuss these under four broad headings:

- probability sampling
- opportunistic sampling
- non-random sampling for representativeness
- theoretical sampling.

The first and third approaches can be seen as concerned with what is sometimes referred to as **empirical generalisation**, whereas the fourth is concerned with **theoretical generalisation** (Hammersley, 1992f). The second approach, on the other hand is not centrally concerned with generalisation at all. Alternatively, the second, third and fourth approaches have been described as **purposive** sampling techniques (Kuzel, 1986).[4]

4.1.1.1 Probability sampling

This is the form of sampling which will be most familiar to researchers from the quantitative tradition. It refers to a method whereby every unit in the universe under study has the same known probability of being studied. The strength of this approach to sampling is that it allows us to use statistical inference to estimate, within precise margins of error, the distribution of a phenomenon of interest in the universe from which the sample has been drawn, when that universe is too large for us to be able to study in its entirety. Given the advantages of probability sampling, in terms of the possibility of generalising from one's findings, it may seem surprising that it is so rarely used in qualitative research. The debate about the possibility of using probability sampling methods in qualitative research can be seen as falling broadly into two camps. On the one hand, there are those who argue that there are significant ways in which random sample is fundamentally inappropriate for the kinds of task in which qualitative researchers are engaged (Lincoln and Guba, 1985; Bryman, 1988; Merriam, 1988; Firestone, 1993). On the other hand, there are those who argue that, while, in principle, there are no objections to using probabilistic sampling methods in qualitative research, it is often impractical or inappropriate (Becker, 1958; Honigman, 1982; LeCompte and Goetz, 1982; Hammersley, 1985; Sandelowski, 1986; Silverman, 1989; Dingwall, 1992; LeCompte and Preissle, 1993a).

Some of those who reject probabilistic sampling methods do so on the grounds that the purpose of such methods is to permit generalisation and that such generalisation is inappropriate to qualitative research (Merriam, 1988; Guba and Lincoln, 1989). Merriam (1988) argued that: "One selects a case study approach because one wishes to understand the particular in depth, not because one wishes to know what is generally true of the many" (173).

[3] The ways in which some writers have sought to redefine the concept of generalisability for qualitative research are discussed in section 5.4; [4] The links between such purposive sampling and Bacon's advocacy of 'tables of discovery' are discussed in section 2.2.1.

Similarly, Stake (1995) suggested that, while some generalisation may be possible from case studies, this is not a primary goal in case study research. The real object of the case study, he argued, is **particularisation** (8). He described how such particularisation is achieved:

> We take a particular case and come to know it well, not primarily as to how it is different from others but what it is, what it does. There is emphasis on uniqueness, and that implies knowledge of others that the case is different from, but the first emphasis is upon understanding the case itself. (8)

The methods of quantitative research, which we take to include random sampling, are seen by Stake as having a potentially distorting effect, insofar as the concern with generalisation leads to a concern to "eliminate the merely situational" (40). By contrast, qualitative researchers are concerned with understanding the individual case in all its uniqueness:

> To sharpen the search for understanding, qualitative researchers perceive what is happening in key episodes or testimonies, represent happenings with their own direct interpretations and stories (i.e. narratives). Qualitative research uses these narratives to optimize the opportunity of the reader to gain an experiential understanding of the case. (40)

From Stake's perspective, when the researcher communicates such idiographic understanding of a single case, this allows the reader to experience the case vicariously (86). The reader is then able to engage in **naturalistic generalisation**, which happens inside the head of the reader, as the (s)he relates the vicarious experience of the case study, with previous knowledge of others settings (Stake, 1994; 1995).

If we accept that the primary function of qualitative research is to provide this kind of vicarious experience and experiential understanding of a particular case, in order to allow the reader to make naturalistic generalisations of the kind advocated by Stake, then probabilistic sampling clearly is both inappropriate and impractical. If one down-plays the goal of public generalisation, the rationale for probability sampling disappears. However, as we shall see in a moment, many qualitative researchers are reluctant to abandon the goal of generalisation in this way. If we adopt Stake's position, we are obliged to see the practice of qualitative research as closer to a communicative art, than a public science. As Kennedy (1979) has wryly remarked, such a position suggests that the researcher needs "the talent of Tolstoi to be able to describe ... events in ways that allow the reader to draw the appropriate inference" (664).

It is also sometimes argued, for example, by Merriam (1988), that probabilistic sampling is of limited usefulness to practitioners, insofar as, by definition, it results in findings which are merely probabilistic. Merriam illustrated her argument in relation to school absenteeism:

> A study might reveal, for example, that absenteeism is highly correlated with poor academic performance – that 80 per cent of students with failing grades are found to be absent more than half the time. If student Alice has been absent for more than half the time, does it also mean that she is failing? There is no way to know without looking at her record. Actually an individual case study of Alice would allow for a much better prediction of her academic performance, for then one would know the particulars important to her situation. (173–4)

The same argument might, of course, be applied to research in healthcare settings. In effect, these authors are arguing that since the purpose of qualitative research is to explain the exceptions that break the probabilistic rule, it is inappropriate to use a sampling method which is designed to produce an estimate of the frequency of the cases to which the rule applies. Firestone (1993) argued that random sampling techniques encourage correlational thinking, which may serve to obscure interactions between variables. This point will be considered more fully when we deal with theoretical sampling in section 4.1.1.4.

Bryman (1988) suggested that one of reasons why qualitative researchers are reluctant to adopt probabilistic sampling techniques is that such techniques are often linked to an emphasis on the individual as the focus of inquiry. Random sampling is associated with what Bryman described as **aggregate psychology** and this is seen as inappropriate to the concern with patterns of relationships and interactions which are often at the heart of qualitative research. Those who reject random sampling as inappropriate for qualitative inquiry may do so on the grounds that it leads to treating a social phenomena as the sum of its parts, rather than grappling with the embeddedness phenomena in their context. Miles and Huberman (1994) expressed this point: "Social processes have a logic and a coherence that random sampling can reduce to uninterpretable sawdust" (27).

However, other authors (e.g. Honigman, 1982; Sandelowski, 1986; Johnson, 1990; Dingwall, 1992; Hammersley, 1992c; LeCompte and Preissle, 1993b; LeCompte and Preissle, 1993c) accept that random sampling could, in principle, be used in qualitative research, but argue that practical and resource

constraints mean that this is rarely appropriate. Dingwall (1992) argued that random sampling techniques, while possible in qualitative research, may be both inefficient and expensive. Honigmann (1982) and Hammersley (1992f) have both suggested that, while statistical inference is the strongest basis for generalisation in qualitative research, it is rarely practical.

A number of problems have been identified in applying probabilistic sampling methods to qualitative research. In particular, the ratio of settings studied to the number in the aggregate is generally too low to allow statistical inference (Firestone, 1993; Hammersley, 1992f). There is concern that, given the small sample size in most qualitative research, random sampling may in fact "deal you a decidedly biased hand" (Miles and Huberman, 1994:27). While there is nothing inherent in the logic of qualitative design which restricts it to small samples, given the intensive and time consuming nature of qualitative research and the associated resource issues, it is rarely possible for participant observation studies to be carried out in more than a small number of settings (Stake, 1994). Similarly, although randomly selected samples of interview respondents, which are sufficiently large to permit statistical analysis, are in principle possible, the time-consuming nature of both the data collection and analysis mean than it is very unusual for sufficient resources to be available to permit the collection of interview data, which would allow statistical inference from sample to population. In both cases, the choice of using non-probability sampling methods can be seen as trading off depth for breadth. The emphasis is upon working with small samples, "nested in their context and studied in depth", unlike quantitative researchers who "aim for larger numbers of context-stripped cases and seek statistical significance" (Miles and Huberman, 1994:27). Stake (1994) has summed up this perspective: "Sometimes it is better to learn a lot from an atypical case than a little from a magnificently typical case" (243).

Various authors have pointed out that it is relatively rare, even in quantitative research, for statistical generalisation to the population of interest to be fully warranted (LeCompte and Goetz, 1982; Bryman, 1988). One of the problems faced by both qualitative and quantitative researchers is that it is unusual for the population of interest to be fully itemised (Firestone, 1993; LeCompte and Preissle, 1993b). Indeed the population of interest may not possess naturally occurring boundaries. While statistical samples may be representative of the particular populations from which they have been drawn, they may not be typical of the population to which generalisation is sought. Honigmann (1982) and Bryman (1988) pointed out that national samples are rare. As a result, while a random sample may be representative of the **sampled** universe, it is rarely representative of the universe to which researchers and readers wish to be able to generalise. This, of course, is a problem for both quantitative and qualitative research.

In summary, qualitative researchers rarely use probabilistic sampling methods. For some this represents a principled objection to their use. Others would be happy to use them, but feel given time and resource constraints, and the opportunity costs involved, conclude that their use is rarely justified in qualitative research. The alternatives to probability sampling are discussed below.

4.1.1.2 Opportunistic sampling

As Hammersley and Atkinson (1995) have remarked, it is not unusual in qualitative research, for opportunism to guide the choice of a group or setting to study. This is often the case with professionals, such as doctors and nurses, who wish to carry out research on their own practice. Such professionals may find themselves in the same position as Pollard (1995) who reflected: "Because of my circumstances my choice reduced to a straightforward decision between doing my research at the school at which I worked or abandoning my desire to do an ethnographic study" (218).

Hammersley and Atkinson (1995) pointed out that pragmatic considerations should not be lightly dismissed in choosing a setting or group for study. They argued that, very often, a wide range of settings will be potentially relevant to a particular research problem and that, therefore, it makes sense to consider practical issues such as ease of access (both geographical and interpersonal) and the costs involved in studying one setting as opposed to another (e.g. travel costs). In some cases, particularly where the phenomenon of interest is highly sensitive, or illicit, the population of interest is highly mobile, or the group of interest is known to be reluctant to participate in research (e.g. Bowler, 1997), opportunistic sampling may be the only avenue open to a researcher.

Honigmann (1982) defined opportunistic sampling as sampling which "follows no strict, logical plan" (81). He argued that such opportunistic sampling has little value until the findings are set in context. He contrasted opportunistic sampling with **judgement sampling**, where the researcher deliberately sets out to sample according to some theoretical

or logical principle. As such, judgement sampling is similar to theoretical sampling, which will be discussed in section 4.1.1.4.

Stake (1994) argued that the central selection criterion in qualitative research should be opportunity to learn:

> My choice would be to take the case from which we feel we can learn the most. That may mean taking the one that we can spend the most time with. Potential for learning is a different and sometimes superior criterion to representativeness ... even for collective case studies (i.e. multi-site case studies), selection by sampling of attributes should not be the highest priority. Balance and variety are important; opportunity to learn is more important. (243–4)

Earlier in this section, we outlined Denzin's position that any instance of social interaction "represents a slice form the life world (Denzin, 1983:133–4). If we adopt the kind of position outlined by Denzin (1983) or Stake (1994), then there can be little objection to opportunistic sampling. Indeed, such opportunistic approaches may be well justified in initial, exploratory qualitative research. If, however, we are committed to strive for cumulative knowledge about the social world and reject the notion that knowledge is simply a mosaic of multiple slices of life, which are to be fitted together to create a whole, opportunistic sampling will be seen as the method of last resort in anything other than the most exploratory social research.

This is not to suggest that pragmatic considerations should not play a part in sampling decisions in qualitative research, as they frequently do in quantitative research. However, elegant the sampling design, it will be useless if limited research funds make it impractical or if it is impossible to gain access to the settings or groups selected. Rather we are arguing that, in selection decisions, such pragmatic considerations should be integrated with a commitment to drawing samples in a systematic and principled way (Schofield, 1993). We now move on to discuss two such principles which underpin sampling decisions in qualitative research.

4.1.1.3 Non-probabilistic sampling for generalisation

A number of authors have pointed out that random sampling is not the only means by which the generalisability of research can be fostered (Delamont and Hamilton, 1976; Kennedy, 1979; Gubrium *et al,* 1982; LeCompte and Goetz, 1982; Silverman, 1989; Hammersley, 1992e; Firestone, 1993): "We must not confuse probability methods with the goal of making claims about representativeness or typicality ... being unable to use probability methods does not rule out the possibility of making reasonable judgements about the representativeness of findings drawn from a particular setting to some wider populations" (Hammersley, 1992f:88).

As Hammersley (1992f) emphasised, a pre-requisite for any attempt at empirical generalisation (whether based on a probabilistic or purposive selection strategy) is that the population to which generalisation is to be made is adequately defined: "Empirical generalisations can only be to finite populations (though these do not necessarily have to be specified very precisely)" (87).

Empirical generalisations may sometimes emphasise typicality. Schofield (1993) argued that, if policy makers wish to evaluate a programme, innovation or intervention, it may be very useful to them to know how it functions in the typical, common, or the ordinary situation. In such cases, selection strategies will be concerned to identify settings or groups which are believed to be typical, in some way, of the aggregate to which generalisation is sought. For example, in studying the process of introducing an innovative treatment, researchers might seek to study it in a hospital department which is believed to be typical in significant respects of the kinds of settings in which it would be applied. Thus they might avoid studying it in state of the art centres or indeed centres known to be particularly poorly resourced or staffed. On other occasions, researchers may deliberately set out to identify atypical settings or groups. For example, researchers may choose to study an unusually successful programme in order to identify the characteristics of settings which enhance such success. In either case selection decisions will be based on a clear identification of the relevant features of the settings of interest (LeCompte and Goetz, 1982).

Delamont and Hamilton (1976) argued that qualitative observational studies[5] may be generalisable insofar as the settings to which generalisation is sought share the characteristics of the setting which is being studied:

[5] See section 4.2.

> It is often argued against anthropological studies that their results cannot be generalized to other settings. This criticism only refers to statistical generalisation. To an anthropological researcher, the development of generally or universally applicable statements is quite a different task, one that is never achieved merely by carrying out a survey. Despite their diversity, individual classrooms share many characteristics. Through the detailed study of one particular context it is still possible to clarify relationships, pinpoint critical processes and identify common phenomena. Later, abstracted summaries and general concepts can be formulated, which may, upon further investigation, be found to be germane to a wider variety of settings. Case studies, therefore are not necessarily restricted in scope. (13)

Similarly, many wards, clinics and other healthcare settings may be expected to share many characteristics. Nevertheless, as Schofield has pointed out, choosing a setting on the basis of its typicality is not a 'quick fix' for the issue of generalisability, since one which is typical on one dimension may not be typical on another.

Hammersley (1992f) suggested that there are at least three ways in which the typicality of findings from particular case studies can be judged. These, in turn, suggest selection strategies which may be employed to optimise the representativeness of findings from qualitative research:

- Where available, published statistics about the aggregate or population to which one wishes to generalise may be used to inform selection decisions. For example, if one wished to study the introduction of minor surgery into general practices, one might start by looking at the national general practices across variables which are judged likely to have relevance for such innovations. Clearly, a decision about which factors are likely to be relevant is a matter of judgement, but they might include the number of partners in the practice, whether the practice is housed in purpose-built or converted premises, the availability of non-medical practice staff, proximity of practices to other surgical facilities, the age of the principals and so on. Decisions about what represents a 'typical' practice would be informed by published and unpublished data available from a number of sources including the Royal College of General Practitioners and health authorities.
- Hammersley also suggested that the combination of case studies and survey research in the same investigation may be useful in increasing the generalisability of qualitative research. Where, the survey element of such studies comes first, it may be used to inform selection decisions in the later qualitative phase. Thus informants or settings for the qualitative phase may be chosen for their representativeness in terms of a central concern of the study, using data collected during the quantitative study.
- Where multi-site studies are possible, or where multiple informants are involved, the researchers may seek to use some form of stratified or quota sampling to ensure that the sample covers the range of variation which is found in the aggregate to which generalisation is sought.

Hammersley (1993f) concluded:

> Empirical generalisation does provide a sound basis for claims to the general relevance of ethnographic studies in the case of both descriptions and explanations. Its use does not necessarily require the study of a large proportion of instances in an aggregate, or a sample selected on the basis of statistical sampling theory. It does, however, require ethnographers to make rational decisions about the population to which generalisation is to be made, and to collect and present evidence about the likely typicality of the case(s) they study. (93)

Hammersley is concerned here with what other authors have called case-to-case transfer (Lincoln and Guba, 1985; Firestone, 1993). In other words, in selecting a particular setting for study, the researcher pays attention to the likelihood that the particular case studied is sufficiently similar to cases to which (s)he will wish to generalise, to make such generalisations defensible. Not all of those who support generalisation on the basis of such case-to-case transfer accept that this should inform sampling decisions (Lincoln and Guba, 1985). This is because they invest the responsibility for making judgements about such transferability of findings in the reader of the research, rather than in the researcher him or herself. These are not judgements which can be made in advance. Rather they depend upon the reader's intimate knowledge of the situation to which transferability is sought and upon the researchers detailed description of the context in which the findings were generated.[6]

4.1.1.4 Theoretical sampling

The fourth broad approach to sampling within qualitative research is theoretical sampling

[6] See section 5.4.

(Hammersley, 1985; Marshall, 1985; Atwood and Hinds, 1986; Bryman, 1988; Merriam, 1988; Hammersley, 1989; Strauss and Corbin, 1990; Firestone, 1993; Henwood and Pidgeon, 1993; LeCompte and Preissle, 1993a; Miles and Huberman, 1994; Keen and Packwood, 1995). Emerson (1981) defined theoretical sampling as sampling: "... in which new observations are selected to pursue analytically relevant distinctions rather than to establish the frequency or distribution of phenomena" (360).

A number of authors (Mitchell, 1983; Bryman, 1988; Silverman, 1989) have argued that it is inappropriate to consider selection issues in relation to case studies in terms of generalisation from samples to populations or universes. Rather sampling should be directed towards making possible generalisations to theoretical propositions. Yin used the term analytic generalisations to distinguish these from statistical generalisations: "In analytic generalization, the investigator is striving to generalize a particular set of results to a broader theory" (Yin, 1989:44 cited in Firestone, 1993:17).

Silverman (1989) argued for the use of theoretical sampling in case study research on the grounds that the aim of such research is to exhibit or test some identified theoretical principle, rather than to achieve representativeness. For him, case study research is dependent upon the adequacy of underlying theory rather than upon empirical generalisation of the kind discussed above. The aim of the researcher is to specify necessary conditions for a phenomenon from among a set of theoretically significant elements.

Silverman (1989) drew heavily upon Mitchell's discussion of the logic of case studies (Mitchell, 1983). Mitchell, unlike Hammersley (1992e) discussed above, argued that **empirical** generalisation from case studies is impossible. He suggests that much confusion has been created by researchers who failed to grasp the distinction between statistical inference and theoretical inference. For Mitchell, the role of case studies is to develop theories: "The documentation of some particular phenomenon or set of events which has been assembled with the explicit end in view of drawing theoretical conclusions from it" (191).

Rather than being concerned, as in probability sampling, with typicality and inferences about the distribution of a characteristic or concomitant variation, the case study researcher should be seeking to make generalisations beyond the sample which are based on the essential linkage between two or more characteristics in terms of some systematic explanatory schema. The basis for generalisation in qualitative research is therefore not representativeness but rather that the case is, in some way, held to either exhibit or test some theoretical principle (Silverman, 1989). In Miles and Huberman's terms, selection in qualitative research should be 'theory driven' (Miles and Huberman, 1994:27).

Firestone (1993) outlined the logic of generalising to a theory, using theoretical sampling. The researcher uses theory to make predictions and then seeks cases that will allow him or her to test the robustness of such predictions under different conditions. There are a number of ways in which theoretical sampling is pursued: "Sometimes analytic generalization attempts to show that a theory holds broadly across a wide variety of circumstances, but sometimes it identifies the scope of a theory – that is the circumstances under which it applies" (17).

In seeking such analytic generalisations, the researcher uses a range of sampling techniques, including the study of negative, critical, discrepant and deviant cases to explore and extend existing theory (Gubrium *et al,* 1982; Firestone, 1993; LeCompte and Preissle, 1993b).[7]

Bryman (1988) took an example discussed by Glaser and Strauss (1967) to illustrate the application of theoretical sampling in a study of the care of the dying in hospital. Glaser and Strauss had developed the hypothesis that the higher the social value of a dying patient (i.e. the higher the nurse's estimate of the degree of impact a patient's death was likely to have on family and/or occupation), the less delay they would experience in receiving services from experts. Bryman commented: "The issue of whether the particular hospital studied is 'typical' is not the critical issue; what is important is whether the experiences of dying patients are typical of the broad class of phenomena (the differential social evaluation of persons and its effects on the delivery of services) to which the theory refers" (91).

The implication for sampling is that, rather than seeking to establish an empirical generalisation by, for example, studying a random sample of

[7] See section 5.4.4.

hospitals in the UK in which people die, one would seek to test the limits of the theoretical proposition (that differential social evaluation impacts upon the delivery of services) in a number of other settings (e.g. outpatient clinics, accident and emergency (A&E) units, general practice surgeries.

Where qualitative researchers are, as Mitchell (1983) argued they should be, seeking to develop theoretical generalisations, this has significant implications for their selection decisions, both initially and recurrently throughout their research. Initially, the researcher may select a case or case(s) for study in relation to their potential for extending or testing existing knowledge. Hammersley (1985) argued that the principle guiding such initial decisions are that the cases chosen should be crucial for theory, in some way. Miles and Huberman (1994) referred to this as the use of "theory driven" sampling (27). An example of such theory driven sampling would be Bloor's work on medical decision making, which was theoretically grounded in Schutz's distinction between habitual and considered interpretations (Bloor, 1994).[8] Bloor pointed out that much of the effort which has been expended upon improving the quality of medical decision making has been predicated on the assumption that medical decisions involve a series of steps, including the accumulation of evidence from investigative procedures, weighing of such evidence, selecting a diagnosis, consideration of treatment alternatives and the eventual selection of treatment. Bloor used the term **polythetic** to refer to this step-by-step process of decision making.

He pointed out that the assumption that medical decision making is polythetic rather than monothetic (i.e. involving a single flash of cognition) has been largely unexamined. If such decision making is largely monothetic and habitual, rather than polythetic and considered, then the implications for strategies and programmes which are designed to improve the quality of medical decision making would be substantial.

The selection of death certification as a context in which to study medical decision making was theory-driven, in that Bloor identified the practice of death certification as one in which non-polythetic decision-making processes were particularly likely to be observed. Unlike many other medical decisions, in death certification, the doctor is solely concerned with diagnosis, since intervention is no longer an issue. In addition, death certification is an unevenly distributed activity, with small numbers of doctors signing a high proportion of certificates. Bloor selected a sample of such high certifiers for his study. This sampling decision was again theory driven insofar as it was informed by the theoretical assumption that repeated exposure to the same object may lead to non-polythetic, routinised decision making.

Similarly, when Dingwall and Murray (1983)[9] chose to focus their analysis upon the way in which children were categorised in A&E units, they were mindful of earlier work on patient categorisation in such departments, which suggested that professionals discriminated between good and bad patients and that categorisation as a 'bad' patient had negative implications for the way in which that patient was managed (Jeffery, 1979). By selecting children as the focus of their study, Dingwall and Murray were able to test and modify the theoretical conclusions of this earlier work.

Again, Silverman (1993) described how his decision to study the ceremonial orders of NHS and private oncology clinics (Silverman, 1984) was conceived as a test of Strong's work on such ceremonial orders in other medical contexts (Strong, 1979a). Using theoretical sampling, the theory which is to be developed or tested is not always drawn from the same substantive area as the proposed study. For example, the study of the general practitioner's (GP's) role in giving lifestyle advice, reported by Parry and Pill (1994), built upon theory developed by Basil Bernstein in the context of education, rather than earlier studies of medical settings.

Finally, Buckholdt and Gubrium's study of *Cedarview,* a residential treatment centre for emotionally disturbed children, illustrates the interplay of theoretical and practical considerations in theoretical sampling (Buckholdt and Gubrium, 1979).[10] The focus of their ongoing research programme was "how people see and understand their lives in institutions whose purpose is care and treatment" (249). Earlier work by Gubrium (1975; 1978), had shown how senility offered a useful/required way of reading elderly people's behaviour in nursing homes, as well as a description of their mental status. Buckholdt and Gubrium chose to build upon this research by studying the treatment practices of care staff and the way in which they are understood and

[8] See appendix 1 for details of this study; [9] See appendix 1 for details of this study; [10] See appendix 1 for details of this study.

organised by them. They described in some detail the way in which they selected *Cedarview* as the setting for such a study:

> Required was a human service institutions with an official goal of treating clients for their problems. This eliminated the kind of services that primarily involved client placement and transfer, the function of many public welfare agencies. We needed a field site where treatment was conducted, preferably on a full-time basis with full-time clients. Nursing homes, mental hospitals, and residential treatment centers met this requirement. Since our focus was to be treatment practice, it was important that the field site selected not only have treatment as a goal but have an ongoing treatment program. Moreover, it was important to select the best institution of its kind (based on conventional criteria) in order to minimize questions that might arise over our interpretations suggesting our findings were not general ones but the result of professional staff members' shortcomings. Again Cedarview fulfilled this requirement, according to public welfare officials and other human service pundits.
>
> Also on theoretical grounds, we had come to be increasingly concerned with the life course generality of professional constructions of client care, treatment, improvement, and decline. We were arguing that not only does people's care status flow with their caretakers' practices, but this happens at all points of the life course, that, in effect there is a continual social processing of human development. Consequently, rather than again focussing on professional practice in nursing homes, we felt that we should study and institutions whose population was much younger, specifically an institution treating children.
>
> The convenience of selecting Cedarview as a field site cut across these reasons. It would be easy to retrospectively present the convenience criterion as the thing that tipped our selection to Cedarview after theoretical considerations had been taken into account. But this was not the way things occurred. Indeed, the relation between our convenience and the drift of our arguments was more muddled. Having Cedarview available to us was no small factor in motivating us to argue for a general theory of professional practice in human service institutions.
>
> Cedarview was a convenient field site on three counts. First, one of our former graduate students was a longtime social worker there. ... Secondly Cedarview was geographically convenient. We had no funding that would permit us to do a study elsewhere than locally ... The geographical convenience made it possible to conduct extensive fieldwork. ... Third, having just completed a major field study of social organisation of care in a nursing home, Gubrium was ready for a change of scene. Buckholdt, on the other hand, was eager to try his hand at systematically studying something he had seen many times over but had not investigated in work on the social context of learning in public schools. (Buckholdt and Gubrium, 1979:250–1.)

Just as theoretical sampling may be used to inform the initial sampling decisions, for example, the initial choice of a setting to study, so too it informs ongoing sampling decisions in qualitative research.[11] As case studies progress, one important form of theoretical sampling is that of seeking out critical tests of the general validity of hypotheses (Silverman, 1989; Dingwall, 1992). Here, the researcher deliberately sets out to study cases where it seems least likely that the hypotheses, which have been developed so far in the field work, will hold good. Again, Silverman's work on oncology clinics (Silverman, 1984), in which he extended the initial sample of two NHS clinics, to include a private clinic, is an example of the way in which sampling decisions taken during the project can be driven by a commitment to develop and test emergent theory. Henwood and Pidgeon (1993) argued that qualitative researchers should pay particular attention to sampling in such a way as to maximise the possibility of finding disconfirming evidence as this will aid the generation of conceptually dense theory.

Glaser and Strauss (1967) recommended the use of progressive theoretical sampling in grounded theory research (see also Charmaz, 1983a; Strauss and Corbin, 1990).[12] Their use of the term theoretical sampling was more restrictive than the way in which it is used by many other methodologists and this reflected the particular concerns of grounded theory. Charmaz (1983a) summarised the approach to theoretical sampling in grounded theory:

> Theoretical sampling means sampling aimed toward the development of the emerging theory (Glaser, 1978; Glaser and Strauss, 1967). As researchers analyze their materials and develop theoretical categories, they frequently discover that they need to sample more data to elaborate a category. Because researchers only develop theoretical categories through the analytic process, they do not know in advance what they will be sampling. Thus theoretical sampling differs from the kind of selective initial sampling most qualitative researchers engage in as they set criteria for their research problem (see Schatzman and Strauss 1973).

[11] See section 4.1.2; [12] See section 4.6.1.4 for a discussion of grounded theory; [13] See section 3.2.1.2 for a discussion of whether grounded theory is best understood as an exclusively inductive procedure.

> As an inductive technique,[13] theoretical sampling exemplifies the inductive logic of the grounded theory approach. Since grounded theorists systematically build their theoretical frameworks out of their observations, theoretical sampling is part of the progressive stages of the analysis. (124–5)[14]

4.1.2 Within-case sampling

In qualitative research, sampling decisions may be made at two levels. First, initial decisions have to be made about the group(s) or setting(s) to be studied. Second, decisions may need to be made about who or what to study within those groups or settings – what Hammersley and Atkinson (1995) referred to as "sampling within the case" (45). As they remarked:

> Decisions must be made about where to observe and when, who to talk to and what to ask, as well as what to record and how. In this process we are not only deciding what is and what is not relevant to the case under study but also usually sampling from the data available in the case ... it is important to make the criteria employed as explicit and systematic as possible, so as to try to ensure that the data about the case have been adequately sampled. (46)

Such sampling, throughout the life of the research is sometimes also called **sequential sampling** (Miles and Huberman, 1994) or **recursive sampling** (LeCompte and Preissle, 1993b). Hammersley (1985) criticised much qualitative research for its failure to sample adequately within cases, in this way, and Goodwin and Goodwin (1984b) presented such within case sampling as an important strategy in achieving **content validity** in qualitative research.

Hammersley and Atkinson (1995) identified three major dimensions along which such within-case sampling occurs.[15] These are time, people and context sampling. In discussing time sampling, they took the hypothetical example of a study of a casualty department in an urban general hospital:

> Any systematic study here would almost certainly reveal different patterns of work and activity according to the time of day or night, and according to the day of the week. The nature of the referrals and emergency presentations would vary too. Saturday nights would probably be characterized by very different rates and patterns of admission from Sunday nights, and so on. Time in our casualty department would also relate to changing shifts of nursing staff, rotations among junior doctors, and so forth. (47–8)

In seeking to adequately represent the people involved in a case, the researcher will be concerned with adequately representing the heterogeneity of such people. In some cases, such people sampling within the case will be informed by demographic criteria. However, in qualitative research, the range of categories of people to be identified will also reflect both member-identified categories and observer-identified categories (Lofland, 1976). Member-identified categories refer to folk categories, which are employed by members themselves. Observer-identified categories are those which are developed inductively by the researcher, on the basis of his or her observations.

The third type of within-case sampling identified by Hammersley and Atkinson (1995) is **context sampling**. As they pointed out, within any setting the researcher is likely to find a number of quite different contexts and these different contexts may be associated with different kinds of behaviour. In seeking to develop an adequate of account of a particular setting, the researcher needs to systematically sample the range of contexts which exist within that setting.

4.1.2 Summary and conclusion

The scope for generalisation from qualitative studies has been widely discussed, particularly in recent years. Where, as is increasingly the case, generalisation is accepted as a legitimate goal for qualitative research, the nature of such generalisation has been contested. We have distinguished

---

[14] Charmaz restricted the term 'theoretical sampling' to sampling which takes place **during** a research project, where the theory involved has emerged **inductively** from the data collected during the project (see also Sandelowski, 1986). As such she was concerned with within-case sampling (see section 4.1.3). Charmaz would exclude from the term theoretical sampling the use of up-front theory to inform either initial sampling decisions or sequential sampling decisions. As Duffy (1985) observed, sampling in grounded theory rejects the use of any preconceived theoretical framework and the theory which informs sampling decisions is expected to arise inductively from the research itself. The tension between this restricted use of theoretical sampling by Glaser, Strauss, Charmaz, Duffy and others and the more extended use of the term by other researchers has led to some confusion. However, most authors now use the term to encompass both activities; [15] Hammersley and Atkinson appear to be primarily concerned here with the possibility of empirical generalisation – in this case with generalising from what is observed and recorded to the setting as a whole. The goal of such sampling is to ensure "as full and representative a range of coverage as possible" (49). However, as the quotation from Charmaz, in the previous section makes clear, such within-case sampling is also a feature of theoretical sampling.

between those researchers who seek to build empirical generalisations and those who strive for theoretical generalisation. The implications of each approach for selection strategies within qualitative research are significantly different. It is important that qualitative researchers should be clear about these differences and that they should be able to justify their sampling decisions in terms of the kind of generalisation they are pursuing in a particular piece of research. In either case, qualitative researchers should be able to demonstrate that their sampling strategies are systematic and logical rather than haphazard.

The goals of empirical and theoretical generalisation are not, however, incompatible. It is possible, in making sampling decisions, to combine a concern with the empirical generalisability of one's findings with a commitment to sampling in ways that are theoretically informed and which seek to test, modify and extend existing theory. In judging the quality of sampling decisions we should be concerned with their fitness for the purpose for which they are designed and systematicity with which they have been carried out.

4.2 Observation[16]

The use of observational research techniques is one of the hallmarks of the qualitative research tradition. Indeed such qualitative research is sometimes referred to as participant observation. Becker (1958) described what is involved in **participant observation**:

> The participant observer gathers data by participating in the daily life of the group or organization he studies. He watches the people he is studying to see what situations they ordinarily meet and how they behave in them. He enters into conversation with some or all of the participants in these situations and discovers their interpretations of the events he has observed. (652)

Unlike more structured forms of observation[17] in qualitative observation, the researchers involve themselves in the settings they study (Goffman, 1961; McCall and Simmons, 1969; Denzin, 1970; Delamont and Hamilton, 1976; Burgess, 1982b; Silverman, 1985; Hammersley and Atkinson, 1995) This interactive nature of participant observation was summarised by Emerson (1981):

> Field research [is] carried out in natural social settings through direct, continuing interaction with people in those settings. Observation through regular **participation** in the naturally occurring activities of the social groupings being studied provides the key characteristic of such research. Such participation both generates the richness of the observation and produces the distinctive methodological and theoretical problems of field research. (351 original emphasis.)

As we shall discuss below,[18] the nature and degree of such participation varies from study to study, and even within the same study at different times (Gold, 1969).

4.2.1 The nature of participant observation

As discussed earlier in this review[19] this approach to studying settings derives in part from the anthropological tradition of studying non-Western societies. Delamont and Hamilton (1976) made this link explicit in relation to educational research:

> Methodologically, 'anthropological' classroom studies are based upon participant observation, during which the observer immerses himself in the 'new culture'. That is, they involve the presence of an observer (or observers) for prolonged periods in a single or small number of classrooms. During that time the observer not only observes, but also talks with participants; significantly, the ethnographer calls them informants, rather than subjects. (12)

In effect, Delamont and Hamilton were suggesting that a classroom or classrooms can be studied in rather the same way as cultural anthropologists studied villages or tribal groups, living alongside the members of the group, interacting with them and in effect learning their culture (Fielding, 1993; see also Burgess, 1982b).[20] Although Delamont and Hamilton focused upon classroom research, the same argument can be applied to settings within health care, such as wards, clinics (Strong, 1979a; Silverman, 1981; Silverman, 1984), A&E departments (Dingwall and Murray, 1983),[21] operating theatres and so on.

Fielding's reference to learning the culture highlights another feature of participant observation. The researcher adopts the role of learner in relation to the setting studied. In keeping with the emphasis in qualitative research on discovery,[22] the participant observer approaches his or her setting with an attitude of humility. Again, drawing

[16] See section 2.5 for an historical overview of the recent application of qualitative observational methods in medical settings in the UK; [17] See section 4.2.2; [18] See section 4.2.3; [19] See chapter 2; [20] See section 2.4.1 for a discussion of the roots of observational research in social anthropology; [21] See appendix 1 for details of this study; [22] See section 3.2.2.5.

upon the anthropological tradition, the participant observer sets out to "appreciate the native's point of view" (Fielding, 1993:156). Fielding illustrates this point with the Native American saying that one should never "criticise a man until you have walked a mile in his moccasins".

Unlike those using structured observational techniques, qualitative researchers are reluctant to impose descriptive categories *a priori* upon their observations (Delamont and Hamilton, 1976). Rather they prefer to generate descriptive categories within the setting they are studying. Where, as increasingly is the case,[23] qualitative researchers seek to build upon earlier studies, they may import some descriptive and analytic categories. However, when they do so, the fit and appropriateness of such categories for the setting under study are made topics for research, rather than being treated as given (e.g. Dingwall and Murray, 1983[24]).

Participant observers generally approach the setting of interest, making the assumption that the common sense notions with which members of the setting operate are likely to be "complex and sophisticated, rather than naive and misguided" (Silverman, 1993:37). Their aim is to "take the viewpoint of those studied, understanding the situated character of interaction" (Silverman, 1993:48). Goffman (1961) explained why participant observation offers particular benefits when the aim of research is to gain an understanding of the **symbolic world** (Denzin, 1970; Silverman, 1985; Fielding, 1993) of members of a setting:

> ... any group of persons – prisoners, primitives, pilots or patients – develop a life of their own that becomes meaningful, reasonable, and normal once you get close to it, and ... a good way to learn about any of these worlds is to submit oneself in the company of the members to the daily round of petty contingencies to which they are subject. (Goffman, 1961:ix–x.)

In particular, the participant observer attempts to discover the "rules which govern the relationships and interactions within the settings studied and to discern the patterns in the behaviour of members of that setting" (Fielding, 1993:157; see also Strong, 1979a). In doing so, participant observers are fundamentally concerned with understanding the routine rather than the exciting or exotic (Silverman, 1993).

Rather than emphasising detachment and objectivity, as in more structured forms of observation, the participant observer chooses to engage closely in the setting being studied. Denzin (1970) emphasised the difference between participant observation and experimental or survey methods, in this respect:

> A central assumption of participant observation is that the investigator shares as intimately as possible in the life and activities of those he is studying. This may involve, as in the case of the ethnologist, moving in with a tribe of Indians and living with them for an extended period of time, or it may consist of joining in the daily rounds of activity of medical students as they attend classes, make diagnoses of patients, conduct laboratory experiments, attend social functions at fraternities, and drink at local bars. In the latter case the investigator may not actually 'live in' with his subjects, but he does partake in as many of their activities as possible. The experimenter or the survey analyst, on the other hand, either brings his subject into his laboratory, or establishes only a fleeting relationship through a structured interview. (187)

In participant observation studies, the researcher typically makes use of a range of methods, rather than restricting data collection to simply recording observations (McCall and Simmons, 1969; Denzin, 1970; Delamont and Hamilton, 1976; Hammersley and Atkinson, 1995). Decisions about how to collect data are guided by pragmatic considerations. The aim is to collect "whatever data are available to throw light on the issues that are the focus of the research" (Hammersley and Atkinson, 1995:1). Denzin (1970) listed the range of methods that may be employed:

> There is then a curious blending of methodological techniques in participant observation: People will be interviewed, documents from the past will be analyzed, census data will be collected, informants will be employed, and direct observation of ongoing events will be taken. For the present purpose **participant observation** shall be defined as a field strategy that simultaneously combines document analysis, respondent and informant interviewing, direct participation and observation, and introspection. (Denzin, 1970:185–6 original emphasis.)

As discussed above, one of the concerns that underpins much qualitative research is a commitment to holism.[25] However, this does not mean that the researcher will necessarily attempt to encompass everything which is going on in a particular setting. As Silverman (1993) argued, the researcher will always enter the field with some perspective or broad focus and it is important to be as explicit as

[23] See section 3.2.1.2 and 3.2.2.5; [24] See appendix 1 for details of this study; [25] See section 3.2.2.3 for a discussion of holism in qualitative research.

possible about this. The research report which is eventually produced will inevitably be a **representation** of the setting study, rather than a **reproduction** of it.[26] As Silverman reminded us: "Contrary to crude empiricists, the facts **never** speak for themselves" (Silverman, 1993:36).

Nevertheless, participant observers normally embark upon their research with a rather broader focus than is typical in quantitative research (Maykut and Morehouse, 1994). For some, such as Becker and Geer (1960), avoiding premature definition of the research problem is seen as crucial to participant observation:

> ... a major part ... of research must consist of finding out what problems he [sic] (the researcher) can best study in this organisation, what hypotheses will be fruitful and worth pursuing, what observations will best serve him as an indicator of the presence of such phenomena as, for example, cohesiveness or deviance. (Becker and Geer, 1960:267.)

As the study progresses, the focus of the research is expected to narrow and data collection to become more targeted. This approach is sometimes known as **progressive focussing** (Silverman, 1993; Hammersley and Atkinson, 1995). Hammersley and Atkinson associated such progressive focussing with the characteristic funnel shape of participant observation studies:

> Over time the research problem needs to be developed or transformed, and eventually its scope is clarified and delimited, and its internal structure explored. In this sense, it is frequently well into the process of inquiry that one discovers what the research is **really** about; and not uncommonly it turns out to be about something rather different from the initial foreshadowed problems. (Hammersley and Atkinson, 1995:206.)

Silverman (1993) cited two of his own studies as examples of the way in which such progressive focussing operates. The first was a study of two NHS cancer clinics. Some way into this study, Silverman refocused his interest upon a comparison of fee-for-service and state-provided medicine as he widened his data collection to include a private cancer clinic (Silverman, 1984).[27] The second was a study of a paediatric cardiology unit, which eventually focused down upon the disposal decisions in relation to a small group of Down's syndrome children who attended the clinic (Silverman, 1981).

One of the strengths of qualitative observation studies lies in their ability to study process rather than merely to record outcome.[28] In educational research, the move towards observational research reflected a dissatisfaction with previous research approaches which largely ignored what took place inside the school or classroom. Delamont and Hamilton (1976) commented on this phenomenon:

> To anyone outside education it may seem paradoxical that such a central area of educational life has previously been a peripheral area for research. But it remains the case, overall, that the classroom has been a 'black box' for researchers, providing merely a vehicle for 'input–output' research designs or a captive audience for psychometric testing programmes. (4)

Similarly Morrison and McIntyre (1969) commented on the lack of attention to classroom processes in educational research up to the 1970s:

> It is almost a cliche of modern educational thinking that pupils' behaviour in the classroom derives largely from their lives outside it. (119)

Delamont and Hamilton (1976) cited Medley and Mitzel (1963:247) as regretting the failure to study educational processes and practices:

> The research worker limits himself to the manipulation or studying of antecedents and consequents ... but never once looks into the classroom to see how the teacher actually teaches or the pupil actually learns. (5)

Much the same could be said about HTA, where, typically, more emphasis is placed upon studying input and output variables rather than studying the practices of health professionals or the interactional encounters between professionals and their patients/clients. As in educational research, observational studies in healthcare settings, have the potential for studying what takes place within the 'black box'.

It is also argued that participant observation studies can give access to important facets of social reality that would otherwise be inaccessible (Wilson, 1977; Silverman, 1993). Silverman (1985) used the example of Becker's study of how someone becomes a marijuana user (Becker, 1953) to illustrate how observation can enable the researcher to obtain data and understandings that would be inaccessible using quantitative

[26] See section 3.2.1.1 for a discussion of the difference between reproduction and representation; [27] See appendix 1 for details of these studies; [28] See section 3.2.2.4 for a discussion of qualitative research and the study of process.

methods. Becker outlined the various stages through which a novice marijuana user passed on the way to becoming a regular user. These included direct instruction about how to smoke marijuana rather than tobacco and learning to interpret its effects and their significance, learning to enjoy the effects, which initially might appear quite frightening, and resocialisation by fellow users after experiencing difficulties and developing greater appreciation of the drug's effects. Silverman argued that this observational study had distinct advantages over, for example, a survey of attitudes of marijuana users, insofar as it allowed Becker to "take the actor's point of view" and to understand the situated and processual character of the transition to regular use of marijuana.

Silverman (1993) distinguished between two rather different theoretical traditions underlying participant observation studies: the interactionist and the ethnomethodological tradition.[29] Within, the interactionist tradition, participant observers are committed to understanding "the creation and change of symbolic orders via social interaction" (47).[30] The concern of interactionist researchers is to take the viewpoint of the actors in a setting, so as to understand the meaning which action and interaction have to those who are involved in it. Bryman (1988) has characterised this concern with meaning (or "seeing through the eyes of" the people in the setting) as an "express commitment to viewing events, action, norms, values etc. from the perspective of the people who are being studied" (61).

This emphasis upon uncovering actors' meanings in participant observation research has been challenged by a number of writers.[31] Silverman (1993) argued, first, that it can result in research that merely reproduces actors' accounts, rather than engages in analysis. Second, he argued that preoccupation with actors' meanings can lead to neglect of the study of actors' practices. Strong (1979a)[32] accepted that there are indeed difficulties in attempting to study the meaning of actions which are observed, arising from the essential ambiguity of human action. There is no foolproof method of ascribing meaning to observations. As Strong put it:

> How can we be certain that such and such really did mean what I say it meant? This difficulty is heightened when the reader is presented with mere snippets of data, extracted from the context in which they occurred – a procedure which is clearly essential when data are long and monographs are short. (230)[33]

However, Strong was not prepared to abandon the commitment to attempting to understanding what is going on in research settings in terms of the meanings of the actors themselves. He argued that it is clear that, in everyday life, "most action is premised upon the assumption that a correct interpretation of it is both possible and likely. Indeed, action is constructed very precisely so that it can be correctly interpreted by the competent insider" (230). He concluded from this that, if the researcher wishes to understand the meanings which underpin action in a setting, then the most effective means open to him is to "immerse oneself as fully as possible in the situation", as participant observers routinely seek to do.

The second tradition underpinning qualitative observational research, identified by Silverman (1993) is that of ethnomethodology. The emphasis here is focussing upon what is observable (e.g. behaviour) rather than what is not (e.g. motivations or attitudes). Rather than seeking to adopt the native's point of view, the researcher should study the activities in which people are employed and the way in which such activities are locally constituted. Silverman quoted Maynard's statement of this position:

> The question that ethnographers have traditionally asked – 'How do participants see things?' – has meant in practice the presumption that reality lies outside the words spoken in a particular time and place. The ... [alternative] question – 'How do participants do things' – suggests that the micro social order can be appreciated more fully by studying how speech and other face-to-face behaviours constitute reality within actual mundane situations.' (Maynard, 1989 cited in Silverman, 1993.)[34]

4.2.1.1 A comparison of structured and unstructured observation

Emerson (1981) distinguished between participant observation and "fieldwork using preestablished schedules of observational categories". He

[29] See sections 2.4.3 and 2.4.4 for discussions of interactionism and ethnomethodology. Chapter 6 illustrates the different research practices that are associated with these two different tradition in relation to studying the impact of medical information systems; [30] The broader debate about members' meanings is more fully discussed in section 3.2.2.1. Here we focus on its implications for observational research; [31] See section 3.2.2.1; [32] See appendix 1 for details of this study; [33] See section 5.4.2; [34] See the case study in section 6.1, which contrasts two pieces of qualitative research that study medical information systems, one of which focuses upon practices, while other is concerned with uncovering meanings.

identified four basic differences between pre-structured observation and more open field methods, such as participant observation:

> a) Rather than developing relevant categories and analytic distinctions over the course of the study, these [i.e. structured observational] approaches use preestablished observational schedules in order to collect data.
>
> b) The data collected are conceptualized in terms of the "frequency distribution of behaviour events" that are assumed to have objective meaning that can be grasped by the outside observer.
>
> c) Prestructured observation studies employ what is essentially a survey design, with prespecified procedures for when and what to observe or sample.
>
> d) The use of these sorts of procedures [for when and what to observe or sample] narrows and restricts the observer's participation in the setting. In pre-structured observations, field interactions are seen as irrelevant (or even a hazard) to the methodology and findings. Participation is unnecessary because there is no concern with trying to enter the worlds of meaning of those studied. Except for its practical impossibility in most natural settings, concealed observation would be the recommended procedure in pre-structured observation settings. (Emerson, 1981:352.)

Educational research probably represents the context in which structured observational techniques have been most fully applied. Delamont and Hamilton (1976) compared the application of qualitative and quantitative methods in educational settings. They take the example of interaction analysis, a quantitative approach to categorising behaviours, which they identified as being "true to the behavioural core assumptions of American psychology" (6), and which, according to Delamont and Hamilton, "uses an observational system to reduce the stream of [classroom] behaviour to small-scale units suitable for tabulation and computation" (6). Using such systems, the observer classifies the behaviour observed into pre-specified categories such as 'accepts feeling', 'asks questions', 'gives directions', etc. Delamont and Hamilton recognised the strengths of such interaction analysis approaches. They are simple to use, reliable and relatively easy to learn. They make possible the study of large numbers of settings and generate data which can be subjected to statistical analysis. They provide numerical and normative data about average practice and allow extrapolation from samples to populations.

However, Delamont and Hamilton also identified a number of shortcomings associated with interaction analysis.

- Most such analysis systems pay little or no attention to the social, temporal and historical contexts in which the data are collected.[35] The time the observer spends in the research setting is typically short and this can lead to neglect of relevant features of the situation.
- Such analysis systems concentrate upon the overt and observable behaviour of participants.[36] In the course of categorising such behaviour, the observer necessarily imputes motive and intention. So, for example, when behaviour is categorised as 'teacher encourages', the researcher is assigning a motive to the teacher. However, interaction analysis systems do not permit any attempt to discover the participant's self-perceived intention.
- Such systems are concerned only with what can be categorised and measured. However, as Delamont and Hamilton pointed out, the definition of categories and category boundaries is not always straightforward. They pointed to the difficulty of deciding whether a particular piece of behaviour should be assigned to 'using student's idea' or 'accepting student's feeling' as an example of the blurring of category boundaries which can sometimes occur.
- By breaking interactions down into multiple small pieces of action, which must then be linked together to form a set of descriptive concepts, the potential of interaction analysis to go beyond descriptive categorisation to theoretical development is limited.
- The use of pre-specified categories involves the imposition of assumptions upon the data which may impede discovery.[37]
- Finally, the 'arbitrary nature of the boundaries of categories in such systems treats reality as frozen or static and serves to limit the potential of such research'.

In summary, Delamont and Hamilton argued that the claims to objectivity and freedom from bias, which are made by some proponents of interaction analysis, are flawed. They suggested that these approaches "risk furnishing only a partial description" (9), as they fail to integrate both the participants' own accounts and the features

[35] See section 3.2.2.3 for a discussion of context in qualitative research; [36] See discussion of members' meanings in section 3.2.2.1; [37] See discussion of discovery in qualitative research in section 3.2.2.5.

of the setting that provide the context in which the behaviour observed is understandable. Similarly, such systems rarely consider the role played by the observer in producing the behaviour which is recorded, this again leads to an incomplete understanding of the data. Finally, the search for statistical extrapolation may have a distortive effect upon research findings:

> In the interests of objectivity, many interaction analysis research studies feel compelled to survey large numbers of classrooms. It is argued (correctly) that small samples may fail to provide statements relevant to the population at large. Such an approach (even if it can achieve true randomness) may however fail to treat as significant local perturbations or unusual effects. Indeed, despite their potential significance for the classroom or classrooms to which they apply, atypical results are seldom studied in detail. They are ironed out to 'blurred averages' and lost to discussion. (10)[38]

Silverman (1993) identified the avoidance of premature definition of variables as one of the strengths of qualitative observational techniques. The use of pre-structured observational schedules necessitates such early specification. Silverman commented:

> Early 'operational' definitions offer precision at the cost of deflecting attention away from the social processes through which the participants themselves assemble stable features of their social world. So, for instance, the qualitative social scientist may be reluctant to begin by defining, say, 'depression' or 'efficiency'. Instead it may be preferable to examine how, in different contexts, 'depression' and 'efficiency' come to be defined. (36)

Strong (1979a)[39] argued that one of the strengths of qualitative observation is that it retains, in the form of the field notes or recordings the researcher has taken, a source of data which is independent of the analysis. Whereas, in more structured approaches, the data are pre-coded by the observer, and, hence, the relationship between the codes and the behaviour observed is not open to inspection, in qualitative observation, at least insofar as field notes are made available in the report, readers are able to check the inferences made by the research.[40]

Strong (1979a)[41] suggested that it is illegitimate to attempt to quantify social phenomena more precisely than lay persons themselves are able to. He argued that some things, such as objects, can be counted with relative ease and that it is therefore appropriate to do so. However, other concepts, such as feelings are less readily quantifiable and that it is misleading to attempt to do so.

Smith (1975), while advocating the use of structured observational methods, acknowledged that in most structured observation schedules there is some degree of inference, also that they often "grossly ignore the complexity" of interaction (203) and focus upon sequentially occurring interaction rather than simultaneously occurring interaction. He also noted the risk of coding biases and of contamination of coding by associated cues.

The emphasis in structured observation is upon minimising contextual factors. Weick (1968:423, cited in Smith, 1975) summarised this emphasis in structured observational techniques: "The general rule is that context should be used as sparingly as possible [in assigning a behaviour to a category], and the immediate situation should be the sole basis for categorization" (224).

This is a clear difference between qualitative and quantitative observational techniques.[42] The latter seek to treat the context in which observational data are obtained as irrelevant, aiming towards generalisations which hold, irrespective of context. Qualitative approaches, on the other hand, assume that behaviour can only be understood in context and that approaches which seek to exclude contextual factors do violence to social reality.

Smith (1975) identified a number of strengths in qualitative observational methods. He argued that such unstructured observation is more appropriate than structured when "the structure or organization is not known to the researcher beforehand ... particularly in the early stages of research" (230). In particular, "the open-endedness of field research may be considered one of field research's strong points in the study of social situations where little is known beforehand about organization, structure, and other key factors" (230).[43] A further advantage of qualitative techniques is their:

> ... ability to shed light on processes or dynamics of social situations. That is, field research data can be gathered over long periods of time for many variables.

[38] See sections 4.1.1.4 and 5.3.6 for discussions of the significance of negative cases; [39] See appendix 1 for details of this study; [40] See section 5.3.4 for a discussion of the importance of displaying enough data in research reports to allow the reader to check inferences; [41] See appendix 1 for details of this study; [42] See section 3.2.2.3 for a discussion of context in qualitative and quantitative research; [43] See section 3.2.2.5 for a discussion of flexibility as a feature of qualitative research.

> Other methods rarely approach this adaptability to the study of social dynamics. For instance, structured observations may be gathered over long time periods but usually for one or a few variables. (230)[44]

4.2.2 Participant observation and interviewing

There has been considerable discussion about the relative merits of observational research, compared with both survey and less-structured interview techniques (Denzin, 1970; Strong, 1979a; Emerson, 1981; Silverman, 1985; Dingwall, 1997a). These arguments are more fully discussed in section 4.3 and therefore touched on more briefly here. Silverman (1985:15–16) argued that "the advan-tage of observational research is that it is able to produce descriptions and explanations appropriate to the way in which people actually behave". By contrast, he suggested that questionnaire and interview responses "provide idealised accounts of attitudes and behaviours which, because they are rationalisations, have an uncertain relation to actual situations".

Emerson (1981) presented observational research as distinctively different from either structured or unstructured interviewing. While he accepted Cicourel's point that there are close parallels between some of the social processes in interviews and those in field studies (Cicourel, 1964), he insisted that the long-term participation of the researcher in a participant observation studies results in research relations which are qualitatively different.

Denzin (1970), comparing participant observation to surveys and interviews, identified a number of advantages of participant observation:

> First the participant observer is not bound in his field work by pre-judgements about the nature of his problem, by rigid data gathering devices, or by hypotheses, as he may be in the survey. The nature of the survey, when it relies on the logic of preplanned inclusion, is to formulate comparison groups, interview schedules, and hypotheses to be tested before, not during, the research process. The fluid, interactive nature of participant observation allows the observer then to combine the best of the survey method with the best of his more unstructured approach.
>
> Second, the observer can avoid the use of meaningless and irrelevant questions. Third, he is better able to make use of his impressions and reactions during the research process than is the survey analyst. Fourth, the observer is in a position to move behind the public selves of his respondents and penetrate the back regions of interaction – regions rarely open to the interviewer in the survey because of his fleeting relationship with the respondent. Fifth, the observer is well-equipped to link statements and actions of his respondents because he is present in the situations where they interact. ... Sixth, the observer method is one in which the best-equipped person is closest to the data as they are collected. .... Seventh, in participant observation greater use is made of informant data and impressionistic reactions than in the survey. This places the field worker in a better position to impute motives from his observations, to pace his observations at a rate that leads to low levels of refusal, and to incorporate what may have first appeared to be irrelevant data into subsequent analysis. Too often the survey analyst is unable to do such pacing and ongoing analysis. (216)

Dingwall (1997a) has insisted upon the primacy of observational research over interview studies on somewhat different grounds. He has suggested that the relative lack of attention to observational work in recent years has arisen partly from external pressures, particularly on graduate students who, faced with rising demands about the amount of training they should undertake during their graduate programmes, and limited resources available to them for carrying out their research, have turned to interview methods, rather than observation, as a way of conducting doctoral research more quickly and cheaply. As discussed in section 4.3, Dingwall and others have argued that non-observational methods: "... generate problems of validity and reliability which are so fundamental that the neglect of observation, and its proxies in direct audio and video recording, fatally undermines many of the conclusions which are alleged to have been drawn" (Dingwall, 1997a:55).

Strong (1979a)[45] identified a number of advantages which observational data have when compared with interview data. In particular, he suggested that interview responses cannot be treated as a guide to actual behaviour. Drawing on Stimson and Webb's study of consultations in general practice (Stimson and Webb, 1975),[46] he pointed out that there is no necessary relationship between what people do in medical consultations and what they say they do when they are interviewed. Strong argued that

[44] See section 3.2.2.4 for a discussion of process in qualitative research; [45] See appendix 1 for details of this study; [46] See appendix 1 for details of this study.

what people say depends upon what questions are asked, by whom and in what context. Again, Strong held that interview responses may not take into account the fact that many people are incapable of immediate recall of many aspects of their behaviour. He reminded us that we fail to notice much of what we do and that this means that interview data have a strong bias against the routine and non-eventful aspects of behaviour, the very mundane happenings, which observational studies are well equipped to capture.

Strong pointed to a paradox in relation to the usefulness of observational and interview data, particularly in a study such as his own, where the focus is upon uncovering the social rules that underpin interactions between professionals and parents of children in paediatric clinics:

> Interviewing generates spoken or written statements of rules, either made by or agreed with by the respondents. These can be openly displayed and sometimes even counted by the researcher. However, there is no way of knowing how these rules relate to and are used in actual behaviour. By contrast, the observation and recording of relatively trouble-free interaction provides extremely rich data on 'what happens', but few if any occasions on which we may actually see or hear the rules being followed. (232)

In other words, interviewing may generate copious amounts of data that are relevant to the researcher's central interest, but there is no way of judging their validity. When the researcher relies upon observation, on the other hand, it may be much more difficult to identify the rules which are being followed, since participants will rarely make these explicit. This means that the researcher is forced to rely upon drawing inferences from behaviour observed about the rules which underlie such behaviour. Since such inferences are always potentially ambiguous, the analyst is obliged to observe a wide range of situations in which the inferred rule would be expected to apply, testing out the inferences in critical cases and refining his or her understanding of the rule and its application. This process is more fully described in sections 4.1, 4.6.1 and 5.4. Clearly, it is more demanding and time consuming than simply asking respondents what rule underlies their behaviour, but is also likely to be more valid.

4.2.3 The role played by the observer in observational research

The central importance of the observer in observational research has occasioned considerable discussion the various roles which could and should be played by such observers. One of the earliest attempts to develop a classification of the available roles was that of Gold (1969), who acknowledged his own debt to Junker (1952). Gold's typology has formed the basis for number of subsequent discussions of roles in participant observation (e.g. Gussow, 1964; Schatzman and Strauss, 1973; Emerson, 1981; Burgess, 1982b; Fielding, 1982). Gold prefaced his discussion of the available roles and the implications of adopting each, with a remark about the significance of whatever role is chosen: "Every field work role is at once a social interaction device for securing information for scientific purposes and a set of behaviours in which an observer's self is involved" (31).

Gold identified four **master-roles** which the observer may adopt. These are **complete participant**, **participant as observer**, **observer as participant** and **complete observer**. The first and the last are the most clear-cut. The complete participant hides his true identity and purpose from those (s)he is observing. For example, the researcher may become a cleaner on a hospital ward, or a counter assistant in a pharmacy, without declaring the intention to use this role to carry out research. As Gold pointed out, the complete participant faces a particular set of personal and situational demands which arise from the constant obligation to pretend to be something (s)he is not. In addition, the ethical issues raised by such covert research are considerable. As Gold remarked, the researcher may: "...suffer severe qualms about his mandate to get information in a role where he pretends to be a colleague in moral, as well as in other social, respects" (34).

The ethical issues raised by the use of covert methods are discussed in section 4.6. Additionally the risk of going native[47] is particularly marked in relation to the role of complete participant.

The complete observer role removes the observer entirely from participation in the setting under study or interaction with the people in whom (s)he is interested. Gold implied that where the researcher adopts this role, the other members of the setting will be unaware of the researcher's presence. However, in structured observations such as those discussed above, the researcher's presence and intention may be known to those observed, but the researcher will avoid any interaction, in the

[47] See section 4.2.4 for a discussion of going native.

interests of objectivity and detachment. In either case, Gold observed, the risk of going native is greatly reduced, but this may be at the risk of misunderstanding what has been observed.[48]

Between these two extremes of complete participant and complete observer, Gold placed the roles of participant as observer and observer as participant. In the participant as observer role, both the researcher and those being observed are aware of the researcher's interest. Gold pointed out that, in the early stages of the research, this awareness of the researcher's purpose may lead to uneasiness and wariness on the part of those being observed. As the researcher's presence becomes more familiar, this uneasiness is likely to disappear. However, as Gold observed, while such a breaking down of barriers may seem to enhance data collection opportunities, there are problems associated with this breaking down of field roles. Gold detailed these:

> Should field worker and informant begin to interact in much the same way as ordinary friends, they tend to jeopardize their field roles in at least two important ways. First, the informant may become too identified with the field worker to continue functioning as merely an informant. In this event the informant becomes too much of an observer. Second, the field worker may over-identify with the informant and start to lose his research perspective by going native. Should this occur the field worker may still continue going through the motions of observing, but he is only pretending. (35)

Gold argued that the researcher who adopts the role of participant as observer must strive to combine a degree of intimacy in the content of the interactions with the informants, while still retaining "sufficient elements of the stranger to avoid actually reaching intimate form" (35).

The final role, identified by Gold, that of observer as participant, involves more formal observation than the previous one. The contact between researcher and researched is also likely to be briefer. While this role reduces the risk of going native, the risk of misunderstanding is increased. As Gold remarked:

> Brief relationships with numerous informants expose an observer as participant to many inadequately understood universes of discourse that he cannot take time to master. These frustratingly brief encounters with informants also contribute to mistaken perceptions which set up communication barriers the field worker may not even be aware of until too late. (36)

Rather than advocating one of these four roles, Gold emphasised the potential and limitations of each. He argued that each role should be seen as "an expedient device for securing a given level of information" (38) and the selection of roles should be governed by considerations of how best those aspects of society in which the researcher is interested can best be studied.

The extent to which the researcher can or should aspire to becoming an insider within the setting studied is implicit in Gold's discussion. Gans (1982) cited Hughes as characterising participant observation in terms of the tension between insider and outsider roles: "The unending dialectic between the role of member (participant) and stranger (observer and reporter) is essential to the very concept of field work and this all participant observers have in common: they must develop a dialective relationships between being researchers and being participants" (54).

Gans identified three possible roles for the observer: total participant, researcher participant and total researcher, but he argued that adopting the role of total participant is likely to lead to the most fruitful kind of fieldwork:

> For only by being completely immersed in an event as an involved person can one really confront and grasp the social and emotional incentives and pressures that act upon people in groups. (54)

This idea that it is by **immersion** in the setting under study that the researcher comes to fully understand the meaning of events and relationships in that setting, underpins much discussion of the methods and merits of qualitative observational research (e.g. Strong, 1979a). However, Gans recognised that this role has serious difficulties:

> It is almost impossible for him to be both a total participant and an observer of himself and other people. Sometimes, one can be a total participant for a short time, and thus obtain empathy into the situations and for the people under study in a direct fashion. (54)

Thus, for Gans at least, being a total participant is a temporary role within the setting, which must be combined with other roles if the research is to be successful.

A number of authors, including both Gans and Gold, point to the intra-psychic difficulties, to which the marginality of the participant observer

[48] See section 4.2.4 for a discussion of ethnocentricity.

give rise to in the researcher. For example, Gans describes the turmoil arising from his own experience of marginality in a research study: "Often, I carried on an internal tug of war, to decide how much spontaneous participation was possible without missing something as a researcher, or without endangering the neutrality which the researcher must maintain when he is studying more than one group, so that he does not risk being rejected by opposing groups" (54).

Uncomfortable as this marginality may be, it is this which Gans, and other authors see as vital for the scientific status of observational research. It is this marginality that distinguishes the academic researcher from the newspaper reporter. The researcher can never fully experience the world as if (s)he were a member. In an observational study, the researcher is likely to be under constant pressure, both from others and from within him or herself, to become involved as a member rather than as a researcher. Gans emphasised the importance of the researcher finding places, outside the setting being studied, where (s)he can de-brief and resume his/her own persona, rather than constantly playing the role assigned in the setting:

> Being marginal and neutral is a constant strain, and the participant observer is always tempted to find someone to whom he can talk freely about his problems as a participant observer and about his opinions on community issues. More generally, he wants to be able to act as a real person, because, most of the time, he is playing a role which does not entirely satisfy his personal needs. For this reason, a spouse, friends and colleagues outside the community are very important; the participant observer – or at least this one – must have someone with whom he can talk personally about his work. (56)

Later writers (e.g. Schatzman and Strauss, 1973; Fielding, 1982) have suggested that Gold's typology may serve to obscure some of the more subtle dimensions of the researcher's positioning within the setting(s) studied. Schatzman and Strauss (1973) emphasised different modes of participation in fieldwork situations, rather than mutually exclusive roles. They suggested that researchers should make different tactical choices at different times and occasions. It would be unusual for a researcher to adopt a single role in relation to all the participants in a setting, or indeed throughout the whole period of observation. For example, knowledge that the researcher is carrying out research may be unevenly distributed in the setting and some participants in the setting may know much more about the research than others. For some the researcher's role may indeed approximate to that of total participant, while for others it may be closer to participant as observer. Equally, where the researcher does actively participate in the setting, the kinds of activities in which the researcher actually takes part, will locate him or her in relation to various category and group memberships within the setting. This will have significant effects upon the data obtained. For example, a researcher studying ward processes, who adopts the role of ward cleaner, will have access to some sites within the ward and not to others. Finally, the extent to which researchers consciously adopt the orientation of insider or outsider in the setting studied will vary.

The researcher's capacity for taking on particular roles within a setting also varies. In some of the classic school ethnographies (e.g. Hargreaves, 1967; Lacey, 1970; Ball, 1981), the researchers adopted an established role within the setting by becoming teachers. Participation to this degree, is not always possible, as the researcher may not be equipped to adopt a role within the setting of interest. For example, in a study of surgeons' or anaesthetists' practice, it would be unusual for the observer to be professionally qualified to participate as an anaesthetist or surgeon. Even where the observer is suitably qualified, adopting a specific participant role within a setting may not be desirable since it may lead to the researcher being unhelpfully identified with a sub-group within the setting and may restrict access to data on other groups. It is also true that attempting to combine an occupational role, such as teacher, with that of observer/researcher is likely to cause considerable strain as the researcher, in effect, tries to do two jobs at one time. For these reasons, many of the school ethnographers cited above combined periods of teaching with periods spent overtly researching the settings.

The way in which the researcher may adopt different roles in different parts of the setting under study and at different times is illustrated by Buckholdt and Gubrium's account of their study of a treatment facility for emotionally disturbed children (*Cedarview*; Buckholdt and Gubrium, 1979).[49] In this study, the researchers did not take on an established role within the setting, in the same way as the school ethnographers described

[49] See appendix 1 for details of this study.

above. Rather they adopted a series of roles, as they carried out their observations in the main areas of *Cedarview*'s operations: the classrooms, the cottages (residences), staffings (case conferences on individual children) and counselling sessions. Their detailed description of the roles adopted illustrates the different tactical choices that participant observers may make at different times and in different locales within the setting of interest (Schatzman and Strauss, 1973). As Wax (1971) noted, many of the most useful roles adopted by the researcher in the course of fieldwork may be spontaneously invented by the combined efforts of the researcher and participants in the setting:

> We sat in and observed the various classrooms individually, rotating our presence among them. Usually, we arranged to sit at a small desk or an empty table at the rear or side of the room. We brought paper and pencil with us and occasionally took notes. We willingly assisted teachers whenever asked and frequently initiated assistance. In time, it was not unusual for us to take over a classroom for a teacher who otherwise could not have taken her coffee break because her teaching assistant was absent for the day. When a teacher circulated among her students to help them individually with one exercise or another, we also circulated. We fit [sic] into classrooms more or less in the informal capacity of assistant teachers. ...
>
> The cottages were easy to observe. All kinds of people come and go from the cottages ... Initially we were one of the crowd ... In short order, as we continued to sit in the cottages every early morning and evening, various personal attachments developed. ...
>
> Cottage life spills over into other places besides cottage areas proper. It extends into the gym, the play ground, swimming pools, and trips away from Cedarview. We fit into these activities in the same way as cottage workers fit in, being combination of supervisor and active participant. ...
>
> We attended staffings together. This seemed to pose little difficulty since, like other participants, we were there to listen, observe and take notes.
>
> Most awkward was our presence in individual counselling sessions between social workers and children. ... Some social workers made us feel very comfortable and included us in the sessions. In other sessions we sat passively at the side or rear of small offices. (Buckholdt and Gubrium, 1979:254–6.)

Detailed description of the researcher's role within the setting studied, such as that given by Buckholdt and Gubrium, is likely to raise concerns about the way in which the researcher's presence within the setting is having an impact upon that setting. Participant observation studies have often been criticised as being especially susceptible to reactivity. In fact, as Smith (1975) has suggested, there is good reason to believe that the problems of reactivity may be as great, if not greater, in more structured observational studies. Where the researcher tries to stay detached from any role in the setting, this may pose greater problems and may be more inhibiting, since "our society has no norms for relationships where a non-member is present but non participating" (224). He cited the list of potential consequences when the researcher seeks to adopt the detached role of the structured observer offered by Weick (1968:370–3). These included "heightened paranoia, creation of hostility and uncertainty, and predictable changes in verbal and behavioural content of those observed" (224).

Strong (1979a)[50] suggested that it is important not to over-estimate the reactivity associated with observational studies, particularly where the researcher engages in the setting for an extended period of time. In the context of his own study of paediatric clinics, Strong observed that his own impact upon the setting was relatively minor because the "daily business of life has to get done and in clinics there was no other time and place for it to be done" (229). Similarly, Smith pointed to what Strauss and co-workers (1964) have described as the "absorbing situation strategy": "Where the situation being observed is sufficiently engrossing and demanding of the subject's attention that he at least temporarily forgets the observer's presence" (Smith, 1975:224).

Strong observed that his own impact on the setting was likely to have been minimised by the presence of multiple other audiences to which paediatricians and parents had to attend, including nurses, other doctors, medical students and so on. Hence the researcher was only one small part of the audience in the setting under study.

Traditionally, field researchers have discussed a range of ways in which participant observers might seek to limit reactivity, largely through being as inconspicuous and unobtrusive as possible. However, as Emerson (1981) remarked, more recently the emphasis has shifted from minimising the researcher's effect to becoming aware of and monitoring that effect.[51] Emerson cited Gussow (1964):

[50] See appendix 1 for details of this study; [51] See section 3.2.1.3 for a discussion of naturalism in qualitative research.

> Ordinarily in good fieldwork, researchers are not greatly concerned about whether they have disturbed the natural field or not, provided that they can analyse how they affected it structurally. Indeed, by affecting it, they often get to know better what it is that they are studying. (Gussow, 1994:231 cited in Emerson 1981:365 footnote)

The key point, as Burgess (1982b) emphasised, is not that the researcher avoid disturbing the setting studied, but that (s)he should take their impact upon the data obtained into account in the course of the analysis.

Emerson (1981) also criticised those who argue that the solution to the problems of reactivity lies in the researcher cultivating a close and confiding relationship with those who are being studied. The weakness in this argument, as Emerson pointed out, is that even trusted individuals are shielded from certain kinds of information. Those being studied will make judgements about the kind of information which is appropriate for particular audiences. While developing trusting and even intimate relationships with participants will grant the researcher access to certain kinds of data, it will also systematically exclude him or her from other kinds of information. As Emerson argued, all relationships predicate **situated access** and all data must be understood as being a product of the relationship between the researcher an researched in a particular context.

The reality that all data are a product of such situated access is illustrated by observations about the impact of the researcher's ascribed characteristics (including gender, age and race) upon the data obtained. The impact of sex and gender is widely discussed in the literature (e.g. Emerson, 1981; Morgan, 1981; Oakley, 1981; Easterday *et al,* 1982; Cunningham-Burley, 1984; Scott, 1984; Warren, 1988; McKeganey and Bloor, 1991; Silverman, 1993).

Silverman (1993) used Warren's study to illustrate how the gender of field workers has a crucial effect upon the data obtained in a setting (Warren and Rasmussen, 1977; Warren, 1988). In a study of nudist beaches the authors found that informants said different things to male and female researchers. When approached by someone of same gender they were more likely to discuss sexual interests. When, on the other hand, they were approached by someone of different gender they were more likely to emphasise values such as freedom and naturalism.

As McKeganey and Bloor (1991) pointed out, most of the discussion of the role of gender in observational studies has focused upon the situation of female researchers. For example, Easterday and co-workers (1982) discussed the particular problems associated with being a female researcher, drawing upon 12 studies in which the authors had been involved. They described a number of roles that participants in the field attempted (sometimes successfully) to impose upon female researchers. These included sexual object, gofer, mascot and daughter. They discussed the ways in which such roles constrain and influence the data obtained.

While acknowledging that "cultural expectations about appropriate gender roles had a pervasive and often unappreciated impact on [our] fieldwork relations and on the nature and type of data", McKeganey and Bloor (1991) cautioned against too ready acceptance of gender as an "explanatory catch-all" (196) in understanding roles and relationships in observational research. Rather, they argued, it is important to recognise that other variables such as class and age may be equally important. They criticised Easterday and co-workers (see above) for assuming that the reason why female researchers have fewer problems in negotiating access is that, as women they are seen as "powerless and non-threatening". They argued that it is equally likely that such perceptions relate to the low status of researchers, whether male or female. In addition, McKeganey and Bloor argued that it is misleading to see the influence of gender as non-negotiable. They cited Hunt (1984) as reporting that in a study of police work the author was able to negotiate a role as 'honorary male' and suggest that such negotiation may be much more viable in an observational rather than an interview study. They concluded that there are likely to be boundaries to the negotiability of roles for men or women, but that within these there may be scope for manoeuvre.

4.2.4 Going native and ethnocentrism

The aspiring participant observer is frequently warned against the dual dangers of **going native** on the one hand, and **ethnocentrism**, on the other (Gold, 1969). Bryman (1988) defined going native as when "the researcher loses his or her awareness of being a researcher and is seduced by the participants' perspective" (96). In other words, rather than being a researcher, the observer becomes a full member of the setting. Fielding (1993) identified going native as "the most important problem researchers face in the field" Bryman reproduced Oakley's account of the experience of going native during her observations of antenatal clinics: "... at three forty-five, after two hours of a busy antenatal clinic I too would sigh with the doctors as we jointly peeped into the corridor

and saw, still waiting, another row of abdomens ... Or at two in the morning I wanted someone to get in there quickly and do a forceps delivery so I could (like them) go home to bed" (Oakley, 1984:128).

The line between empathetic understanding and going native is a fine one, as Oakley's observation illustrates. Indeed, one could argue that, as she was able to achieve sufficient distance to reflect upon the implications of her feelings, she had not in fact gone native at all.

The most frequently recommended strategy for avoiding going native is that of taking frequent breaks from the setting in which observation is being carried out (Gold, 1969; Smith, 1975). Such breaks, and in particular discussions about the emerging analysis with mentors and others, offer the researcher the opportunity to re-establish the marginality which is seen by many as essential to the conduct of good fieldwork.

In avoiding the risk of going native, the observer may run the alternative risk of ethnocentrism – that is of remaining so detached from the setting under observation that (s)he fails to penetrate the superficiality of his or her own initial observations of the setting. Both Gold (1969) and Hammersley and Atkinson (1995) have presented this danger as particularly associated with the complete observer role, where the researcher does not involve him or herself in the setting under study. The risk, as Hammersley and Atkinson observed, is that the researcher seriously misunderstands the behaviour which is being observed.

4.2.5 Covert and overt observation

A number of participant observation studies (e.g. Rosenhan, 1973; Holdaway, 1982; Clarke, 1995) have been carried out without the knowledge or consent of those being observed. This has led to considerable debate about both the ethics and the effectiveness of such covert observational studies. The ethical issues raised are discussed in section 4.6. Here we restrict the discussion to the practical and methodological issues raised by covert research.

Such covert research is defended either on the pragmatic grounds that if permission had been sought, it was likely that it would have been refused (e.g. Holdaway, 1982) or that the reactive effects of having a covert observer in the setting would have invalidated the findings obtained (Lofland, 1971, but see above). Hammersley and Atkinson (1995) argued that the assumption that access for overt research will be blocked is sometimes ill-founded. They cited Fielding's study of the National Front, for which he received official permission, as evidence that it is possible to obtain access to apparently unlikely organisations (Fielding, 1982).

Where observers seek to keep their research a secret from participants in a setting, this often causes practical difficulties in carrying out the research. One problem relates to the demands of combining the work of observing and recording with those of holding down a normal role within the setting. For example, Shaffir (1985), described how he experienced difficulty in combining a full-time clerical job within a Jewish Hassidic community with his research. As well as this, covert observation is likely to restrict the access of the researcher to certain parts or groups within the setting under study (Gans, 1982). Where (s)he has gained access by posing as a *bona fide* role member within the setting, the researcher will be restricted to investigating those sites within the setting that are open to such a role member. Thus, for example, a hospital porter, would be unlikely to obtain access to case conferences. Similarly, the people to whom the researcher is able to address questions are likely to be restricted to those with whom such role occupants would normally interact. Burgess (1982b) also suggested that the risk of going native is greater in covert research.[52] In addition, the personal stress experienced by the covert researcher is noted by a number of authors. For example, Gans (1982) described the discomfort of the researcher who is constantly in a situation where "he is always taking and never giving". The covert researcher has to pretend to participate emotionally as a full member would and Gans argued that such psychological dishonesty can lead to a pervasive feeling of guilt. Again, the preoccupation with monitoring one's own role presentation may so distract the covert researcher that (s)he is unable to give adequate attention to the settings and behaviours which are being observed (Cassell and Wax, 1980). There are some, however, such as Hilbert (1980) who have argued that the covert researcher is obliged to learn how to behave as a competent member of the setting and this, in itself, will offer significant insights.

As Hammersley and Atkinson (1995) pointed out, the choice between covert and overt observation is not as clear cut as is sometimes suggested. Indeed it is difficult to imagine a situation in which the observer was entirely open about the exact nature

[52] See section 4.2.4 for a discussion of the risks of going native in observational research.

of his research intentions with all participants in a setting. Even where the researcher is committed to openness, the participants may have neither the interest nor the background knowledge to fully grapple with the research objectives. For example, Buckholdt and Gubrium (1979)[53] described the detailed procedures for negotiating access, which were adopted in their study of a treatment facility for emotionally disturbed children.[54] These procedures included meeting with the members of each department in the facility to presents their plan of work and distributing copies of a description of their project to all staff members. However, in spite of this, they became aware that, for the most part there was a fundamental misunderstanding of the object of their research among the staff in the facility. The researchers described this:

> The feedback we received during and after the presentations to the administrator and the departments made it clear that they knew we would be studying the various professional practices from testing and staffing to teaching and child care. Except on a few well-contained occasions, however, they understood practice to be a matter of application and our study as being done in the service of improving the application of their skills to the treatment, care and evaluation of emotionally disturbed children. Our own understanding of practice, on the other hand, was more basic in that we were, in effect interested in the practice of professional applications of skills. We left the difference in understanding stand, not bothering to develop it for staff members. (253)

4.2.6 Summary

In this section, we have reviewed the place of observational studies within qualitative research and discussed the advantages proposed for such studied, as compared to both more structured forms of observation and interview studies. We have described the various roles that may be adopted by the researcher in observational research, and argued that the role played by the researcher must be taken into account in interpreting the data obtained. We have reviewed literature dealing with the risks of reactivity, over-rapport and ethnocentrism, as well as discussing the advantages and limitation of covert and overt observational methods.

4.3 Interviews

Interviews are widely used in HTA for a wide range of purposes, including discovering how consumers evaluate the services they are offered, what understandings and attitudes underlie particular kinds of health behaviour and what might be required to persuade people to change health-related behaviours. The popularity of interviews as a means of data collection partly reflects the recognition that people do not merely respond to stimuli but act on the basis of their interpretations of the world around them and their experiences within it. Interviews have been used in an attempt to access such interpretations to discover what people think about the world they live in, how they evaluate their experiences within it and why they behave as they do. Put simply, the choice to use interviews to collect data can be interpreted as, "If you want to understand what people do, believe and think, ask them". Standardised interview surveys have long been commonplace in HSR and, increasingly, more qualitative approaches to collecting interview data have been used. While traditionally interviews have been one-to-one encounters between the researcher and one interviewee, recently group interviews (sometimes known as focus groups) have gained popularity in both academic and market research (Eriksson, 1988; Khan *et al,* 1991; Morgan, 1993, 1996; Woodward, 1993; Krueger, 1994; Kitzinger, 1995; Dilorio, 1994; Secker *et al,* 1995; Betts, 1996). While there are significant differences in the practical tasks associated with carrying out such group interviews, the methodological issues raised by single and group interviews are largely the same. We have therefore chosen to treat individual and group interviews together in this review.

In later parts of this section, we shall raise some fundamental problems related to the use of both quantitative and qualitative interviews, but before doing so, we shall outline the different types of interview commonly in use and compare qualitative interviews with both quantitative interviews and participant observation.

4.3.1 Types of interview

It has been commonplace to draw a distinction between **structured** and **unstructured** interviews (Burgess, 1982b; Habermann-Little, 1991; Fontana and Frey, 1994). However, the term 'unstructured' is misleading insofar it is impossible to conceive of an interview which is totally without structure. It is more accurate to see interviews as ranged across continuum of standardisation (Richardson *et al,* 1965). Denzin (1970) identified three types of interview:

[53] See appendix 1 for details of this study; [54] See section 4.2.3 for further details of this study.

- The standard schedule interview "in which the wording and order of all questions is exactly the same for every respondent" (123) and the instrument in administered in the same way to all respondents.
- The non-schedule standardised interview, in which the interviewer "works with a list of the information required from each respondent" (125), (quoting Richardson *et al,* 1965:45), but the particular phrasing and ordering of questions is adapted to suit individual respondents.
- The non-standardised interview, in which no specific set of questions is employed and questions are not asked in any particular order. Such interviews are sometimes described as conversational though, as we shall see, there are problems with this designation.

Denzin argued that standardised non-schedule and non-standardised interviews offer the advantages of allowing respondents to display their ways of understanding the world,[55] recognising that no fixed sequence of questions will be appropriate for all respondents and allowing respondents to raise issues which had not been anticipated by the researcher.[56]

Other authors have used different terms to describe similar points on the continuum from standardised to non-standardised interviews. Britten (1995), for example, distinguished between structured, **semi-structured** and **depth** interviews. McCracken (1988) and Crabtree and Miller (1991) used the term the **long interview** to describe an approach which is similar to Denzin's non-schedule standardised interview.

4.3.2 The critique of the standardised interview

Qualitative interviews are frequently presented as a response to the perceived shortcomings of standardised interviewing. Preoccupation with standardisation is seen as arising from the positivist assumption that interviews are concerned with eliciting a body of facts from respondents which exist "out there in the world" (Silverman, 1985: 160). The aim of standardisation is to isolate such facts from the context of the interview. As Denzin (1970) noted, the logic underlying the standardised interview is that, by making the order and the wording and the presentation of interview questions the same for all respondents, the interviewer can be sure that any differences between respondents which are uncovered are attributable to real differences rather than to differences in the instrument or research procedures. Standardisation is thus seen as the antidote to bias and as a means to isolating respondents' true opinions from the distortion of response effects (Mishler, 1979).

Critics of standardised interviewing have argued that, such interviews do not and cannot achieve the degree of control they aspire to and, therefore, cannot be treated as a gold standard against which qualitative interviews can be measured and found wanting. Mishler (1979) summed up this critique: "The standard survey interview is itself essentially faulty and ... it cannot therefore serve as the ideal methodological model against which to assess other approaches" (29).

Cicourel (1964) argued that the goal of the standardised interview is simply unrealistic. The interview situation is a social encounter between two people and as such it is governed by the same situational factors as other such encounters. While it may be possible to standardise the question wording and the order in which questions are presented, it is not possible to achieve the degree of standardisation of interviewer behaviour which would ensure that the same questions are asked **in the same way** of all respondents. Both the interviewer and the respondent will bring their own stock of knowledge and definitions of the situation to each interview. These are not limited to the topic of the interview, but will include "their attractiveness or unattractiveness to one another, their bodily presence, the social, physical and role distance" (80), and, just as in everyday life, these factors play an important part in the way in which both parties act in the interview situation. As a result, standardised interviews cannot be treated as the uniform presentation of identical stimuli. Cicourel (1964) summed up the problem facing those who advocate the standardisation of interviews in the interests of reducing interviewer bias and increasing reliability: "Comparability is not possible in the sense of the classic experiment of exposing the same conditions to the same sample of subjects in identical fashion with complete controls" (87).

Cicourel (1964) also argued that interviewers are faced with objectives that are often fundamentally incompatible. For example, they are required to

[55] See section 3.2.2.1 for a discussion of meaning in qualitative research; [56] See section 3.2.2.5 for a discussion of flexibility and discovery in qualitative research.

administer the questionnaire in an identical way to all respondents, while at the same time developing rapport with each respondent so as to elicit frank answers. The interviewer is expected to adapt to the respondent while at the same time following the same procedures for every respondent. Once again, Cicourel argued that, it is impossible to ignore the similarity between interviews and everyday social interactions: "Canons of research demand that the interviewer operate somewhat like a computer with all the appearances of a fellow human being, but, so far as we know, persons in everyday life find it impossible either to present themselves as both or to receive presentations of others (regardless of the form it takes) which conform to the strict canons of scientific inquiry" (89–90).

He summed up the dilemma facing the researcher who aims for standardisation of the interviews:

> It is doubtful if the interviewer can be trained to utilize literally the principles of social process in interviewing, for this would mean programing him like a computer ... but we want an interviewer who would be completely flexible with respect to mood, affect, appearance etc., in his presentation of self as an interviewer; all of this while obtaining the standardized information required from a standardized schedule in a way which takes all idiosyncratic, situational, and problematic features into account. If the interviewer were like a robot with built-in speaking and tape recording equipment, this would insure standardization and insure the researcher that standardized stimuli are being presented to the subject, but it would not allow any flexibility in the presentation of self. (90–1)

A further charge levelled at the quest for standardisation in interviews is that it mistakes the standardisation of questions with the standardisation of meaning to the respondents. In responding to questions, respondents are involved in a process of interpretation. In other words, although an identically worded question may be administered to all respondents, this does not guarantee that it has the same meaning for all respondents. Denzin (1970) argued that "individuals have unique ways of defining their world. To understand that world meaningfully, researchers must approach it from the subject's perspective" (125). Similarly, Mishler (1979) argued that the statistical analysis of survey data depends crucially upon the assumption that the meaning of both questions and responses can be treated as identical for all respondents and that this assumption fails to recognise the possibility of variation in meaning across sub-groups of the population.

Voysey (1975)[57] criticised the use of pre-defined interview schedules on the grounds that they take for granted the very features they should be examining and gloss over the unexamined assumptions which are embedded in the process of analysis of such interview data. She argued that:

> Questionnaires in effect filter the social processes under study through a pre-defined 'grid' of categories assumed to represent the range of possible alternative responses appropriate to the area of research. Fixed choice (yes/no) questions represent the extreme in this respect, but scaling techniques may be no less inappropriate. It is meaningless to produce measurements or qualifications of phenomena whose dynamics are not yet understood.
>
> Furthermore the advantages claimed for questionnaires would rarely stand up to scrutiny. They are supposed to eliminate observer biases, to provide a routine methods of investigation and analysis which presents the same stimuli to all respondents. Rather, at every stage, a host of assumptions and interpretations are made by everyone employed on a project, which are commonly unacknowledged and uninvestigated in the presentation of results. (66)

From a feminist perspective, DeVault (1990) argued that distortions and misunderstandings are particularly likely to arise in situations where the researchers' categories and language reflect a male perspective.

Advocates of a less standardised approach, suggest that rather than being preoccupied with standardising the questions as in survey research, researchers should attempt to achieve "equivalence of meaning" (Richardson *et al,* 1965) for all respondents. This would entail encouraging interviewers to tailor their questions to the vocabulary and understandings of individual respondents, rather than imposing a uniform set of stimuli, which may be differently interpreted by different respondents. Put differently, less standardised approaches to interviewing are concerned to fix the meaning rather than the wording of questions and advocate the flexibility to rephrase questions to fit each individual's interpretation (Denzin, 1970). Interviewers and their respondents can then be seen as "talking together", rather than behaving as "stimulus senders and response emitters" (Mishler, 1979:22).

[57] See appendix 1 for details of this study.

The risks of misinterpretation inherent in survey approaches are illustrated in Kirk and Miller's discussion of the way in which the definition of a household as "persons who cook their rice from the same pot" led to highly misleading conclusions about the distribution of wealth in a part of Sri Lanka (Kirk and Miller, 1986). Contrary to the researchers' assumptions, every married woman in Sri Lanka has her own cooking pots and hence a household could contain many cooking-pot units.

Quantitative interviews are also criticised for what Mishler (1979) has described as **context stripping**.[58] He argued that, in everyday conversations, ascertaining the meaning of people's answers to questions is highly dependent upon contextual information. In the absence of such contextual clues, we are vulnerable to misinterpretation of the meaning which others intend to convey in their answers. The procedures of survey interviewing are criticised for systematically stripping away the context of individuals' responses to survey questions, while the techniques for analyses of such interview data represent efforts to ascertain the meaning of responses in the absence of the contextual grounds for understanding, which are present in everyday conversation. Mishler summed up this position: "A meeting between strangers, unfamiliar with each others' 'socially organized contexts' does not provide the necessary contextual basis for adequate interpretation" (24).

Finally, some researchers have criticised standardised approaches to interviewing on political and ethical grounds. Oakley (1981) argued that conventional approaches to interviewing are inappropriate for feminist research, as they are based on a hierarchical, exploitative relationship between researcher and respondent. They involve the manipulation of respondents as sources of data, though this may not always be apparent since the cooperation of respondents depends heavily upon their perception that they are being kindly and sympathetically treated. Such approaches are seen to impose passivity on the respondent and oblige the respondent to mechanically serve the purposes of the researcher. Oakley advocated a more reciprocal relationship between interviewer and respondent in which the interviewer responds willingly to women's questions, requests for information about herself and about the research. Similarly, Stanley and Wise (1983) argued that standardised interviews are politically unacceptable on the grounds that they maintain a hierarchical relationship in research.

4.3.3 Virtues of qualitative interviews

Advocates of qualitative interviewing identify a number of benefits, which they believe to be associated with this approach.

4.3.3.1 Accessing the respondent's definitions and interpretations[59]

First, qualitative interviews are seen as offering the possibility of exploring the way in which respondents themselves define the experiences and practices which are the object of the research (Paget, 1983; Griffith and Smith, 1987; Merriam, 1988; Jensen, 1989; West, 1990; Crabtree and Miller, 1991; de Vries *et al,* 1992; Oiler Boyd, 1993b; Kleinman *et al,* 1994; Britten, 1995; Secker *et al,* 1995). In contrast to quantitative research, which can be seen as pre-defining the topic and imposing a **grid** (Cicourel, 1964) on informant understandings, qualitative approaches to interviewing are believed to offer respondents the opportunity to define the problem in their own terms and to challenge the researcher's pre-conceptions about what is important or significant about the matter at hand.

Crabtree and Miller (1991) listed some of the topics within healthcare settings that they believed would be amenable to investigation using less structured interview methods:

> Discovering patient- and physician-specific models of common illness problems; understanding the connections of smoking drinking and food to the core concerns of patients' lives; defining the meaning of function, social support, quality of life, quality of care and stress from a patient-centred perspective; revealing the effect of diagnostic labels on patients' self-perceptions; discovering the effect of consumerism and a product-based understanding of health care delivery on physicians' self-understanding; describing the content of suffering, empathy, vulnerability and hope; and identifying the content of ethical decisions. (145–6)

In all these cases, qualitative interviewing is advocated as a means of uncovering the "insider's perspective" (Jensen, 1989), "how research participants understand their world" (Secker *et al,* 1995), the "actor's frame of reference", (Oiler Boyd, 1993b) "what is on someone's mind" (Merriam, 1988) or the "interviewee's own framework of

[58] See section 3.2.2.3 for a discussion of context in qualitative research; [59] See section 3.2.2.1 for a critique of 'members' meanings' approaches in qualitative research.

meanings" (Britten, 1995). Patton (1980) argued that interviews are used when the researcher wants to find out something which cannot be directly observed:

> We cannot observe feelings, thoughts and intentions. We cannot observe behaviours that took place at some previous point in time. We cannot observe situations that preclude the presence of an observer. We cannot observe how people have organized the world and the meanings they attach to what goes on in the world – we have to ask people questions about those things. The purpose of interviewing, then, is to allow us to enter into the other person's perspective. (196)

McCracken (1988) described the potential of the qualitative interview in similar vein:

> The method can take us into the mental world of the individual, to glimpse the categories and logic by which he or she sees the world. It can also take us into the life world of the individual, to see the content and pattern of daily experience. The long interview gives us the opportunity to step into the mind of another person, to see and experience the world as they do themselves. (9)

As Silverman (1993) remarked, this approach views respondents as experiencing subjects who actively construct their worlds and the aim of the interview is understood to be that of generating data which grant authentic insights into such experiences. Unstructured, open-ended interview are seen by many as the most appropriate means of achieving such insights. In particular, there is a commitment not simply to **describing** the way in which insiders understand their world, but also to treating any differences between the subject's understandings and those of the researcher as legitimate, cultural differences, rather than seeing them in terms of an unfortunate departure from the norms of the researcher of the interests (s)he represents.

By the skilful use of qualitative interviewing, such researchers aim to avoid imposing their own structures and assumptions upon respondents' view of the world (Britten, 1995).[60] Uncovering the meanings, beliefs, understandings and cultures of the respondents acts to counter researchers' (or policy makers') taken for granted assumptions (Secker *et al,* 1995). Marshall and Rossman (1989) argued that using qualitative interviews, rather than formal, structured interviews, allows the researcher to respect "how the participant frames and structures the responses ... the participant's perspective on the social phenomenon of interest should unfold as the participant views it, not as the researcher views it" (82).

Pill (1995) has argued that using highly structured method of data collection can prevent researchers from uncovering the range and depth of people's feelings and opinions, which may be more accessible through qualitative interviews. She examined a cross-sectional survey, which aimed to "determine whether patients think that it is appropriate for general practitioners to undertake health promotion in the consultation" (Pill, 1995:44 citing Wallace and Haines, 1984). These authors had used a forced choice format to confirm that patients felt that it was appropriate for GPs to undertake health promotion in general practice consultations. Pill argued that by limiting their data collection to the use of standardised questions, the researchers failed to access the complexity of patients' thinking about health promotion in the consultation. She suggested:

> Another option would have been to explore in much more detail exactly what experiences patients had had of health promotion, their attitudes to their general practitioners, and the circumstances under which they felt it was or was not appropriate for the doctor to bring up such topics in the consultation. The aim would be to avoid imposing pre-conceived ideas on the respondent, but instead to draw them out and understand their perspective by using less structured and more open questions. (Pill, 1995:47.)

One example of the way in which less structured approaches to interviewing can give access to respondents' definitions and interpretations is Bloor's recent study of medical decision making, which is discussed in chapter 3 (Bloor, 1994).[61] As Bloor noted, a great deal of effort is currently being expended upon improving the quality of medical decision making. Such efforts are predicated upon the assumption that such decision making involves a series of steps during which doctors weigh possible interpretations, the available evidence and the range of alternatives open to them, before selecting a course of action. Bloor used semi-structured interviews with doctors responsible for death certification to demonstrate that, at least in this case, such taken for granted assumptions about how doctors make decisions are unwarranted. Rather, for many of the doctors he interviewed, death certification had become a routinised activity, which minimised the need for reflection and weighing of alternative possibilities.

[60] See section 3.2.1.2; [61] See appendix 1 for details of this study.

4.3.3.2 Penetrating respondents' public accounts

Some researchers advocate the use of qualitative interviews on the grounds that such methods are more likely to extract a true version of events or experiences from respondents. From this perspective, structured interviews are seen as highly artificial events, which are likely to produce superficial and misleading accounts from respondents. Denzin (1970) argued that there is no *a priori* reason to assume that respondents will be sincere and tell the truth to interviewers. He sees tightly structured interviews as particularly likely to generate misleading accounts. On the other hand, informal, conversational interviews, during which the researcher is able to develop a relationship with the respondents, are seen as more likely to create a context in which respondents are prepared to 'come clean', particularly about sensitive topics. Such interviews enhance the possibility of "penetration into their (respondents') relational worlds" (Denzin, 1970:133).

West (1990) claimed that, by using a combination of observation and qualitative interviewing, in his study of families with an epileptic child, he was able to penetrate beyond the parents' official versions of what it is like to care for a chronically ill child. As a result, he believes that he was able to access the private accounts, which were much gloomier and included trenchant criticism of the health professionals involved in their children's care. West criticised other interview studies, such as that of Voysey (1975),[62] which have reported more positive accounts of the experience of parenting ill or disabled children, for having failed to penetrate the fronts that parents offer in the public accounts.[63]

As Secker and co-workers (1995) have suggested this concern with uncovering true feelings is typical of marketing approaches to qualitative interviews, where there is an assumption that objective knowledge can be achieved through improving the technical aspects of the interview. Interviewers are encouraged to develop maximum rapport, sustain eye contact, exploit body language and so on as ways of improving the quality of the data obtained. They are trained to create a comfortable relationship with their respondents, adopting a non-directive approach, while remaining sensitive to and following up throw away remarks etc.

Silverman (1993) has taken up this claim that depth interviewing creates deep mutual understanding between researcher and respondent. He recognised that this humanistic approach is superficially seductive' especially since: "It seems to blend a self-evident truth about humanity with political correctness about the need for mutual understanding and dialogue" (95).

Likewise, Eakin and Maclean (1980) suggested that the appeal of qualitative methods for health promotion research often reflects the perceived ideological compatibility of such methods with health promotion practice, insofar as they reflect a "belief in the capacity of individuals (non-experts) to generate useful knowledge and insight" (72).

However, Silverman (1993) could see no reason to believe that such interviews necessarily lead to the "exchange of unique human experiences". Just as we are suspicious of the media's claim to access personal experience through interviews with celebrities, we should be wary of the claim that research interviews have uncovered authentic human experience. In both cases we may have done nothing more than elicit familiar and socially acceptable ways of accounting for success or failure.

Similarly, Dingwall (1997a) argued that, however informal, an interview is not the same as a conversation. Its key feature is that respondents are "put on notice to talk about **something**" (58) (original emphasis) and the interviewer defines what the parties will talk about and what is deemed relevant:

> The sequence may be flexible; the question wording may be flexible; it may be dressed up like a conversation between friends. But an interview is not a conversation between friends. It is a deliberately created opportunity to talk about something which the interviewer is interested in and which may or may not be of interest to the respondent. If the interviewer refuses to propose topics, the respondent is obliged to guess what might be relevant until the interviewer gives some indication that they are happy with the line being taken. (59)

4.3.3.3 The flexibility of qualitative interviews

Qualitative interviews are often seen as particularly suitable for exploratory or hypothesis generating types of research.[64] Burgess (1982a) argued that,

[62] See appendix 1 for details of this study; [63] It could be argued that West's criticism of Voysey is misplaced, insofar as her analysis was explicitly **not** concerned with uncovering what the parents of disabled children really thought or felt. Rather she was concerned to analyse the ways in which the parents **presented** themselves and their children and she argued that what parents 'really thought' was inaccessible; [64] See section 3.2.1.2 for a discussion of flexibility as a feature of qualitative research.

unlike standardised interviews, qualitative interviews allow researchers to follow up interesting ideas and to open up new dimensions, which had not been anticipated in advance of the interviews (see also Denzin, 1970). Britten (1995) also identified the flexibility of qualitative interviews as a means by which to "uncover new areas or ideas that were not anticipated at the outset of the research" (252). Similarly, de Vries and co-workers (1992) identified one of the strengths of the qualitative interviews used in their study of a health education programme as their capacity for enhancing the generation of ideas and theories.

The benefits of such flexibility are demonstrated in a qualitative interview study of lay beliefs and responses to antihypertensive therapy (Morgan and Watkins, 1988[65]). The authors found substantial differences in adherence to medication between white and West Indian patients. Whereas compliance was high among the white respondents, less than half of the West Indians were classified as compliers. The flexibility, which is characteristic of qualitative interviews, allowed the interviewers to discover that many of the West Indian patients were regularly 'leaving off' their tablets. This involved either not taking their drugs for a few days each week or not taking them for a week or two, or even a couple of months. Such 'leaving off' was a deliberate action which was taken even though the respondents were aware that their doctors expected them to take their tablets continuously. The researchers were able to exploit the flexibility of the interview situation to explore the lay logic that underpinned the respondents' decisions to act in a way which was contrary to the advice which they received from their doctors.

4.3.4 Validity in qualitative interviewing[66]

Advocates of qualitative interviews have identified a number of threats to the validity of qualitative interviews. Denzin (1970) warned: "The investigator must show the extent to which his questions measure what is intended, as well as demonstrate the reliability of his instrument" (132).

Denzin drew attention to potential sources of invalidity in interviews, including the possibility that the information derived from interviews will be distorted by interviewers making the unwarranted assumption that they share a common perspective with their respondents, as well as problems related to the role the interviewer adopts in the interview, the transience of the research relationship which means that the respondents may fabricate their responses, status inequalities between interviewer and respondent, and the constraint of particular interview contexts. As Silverman (1985) pointed out, there is some inconsistency between this concern with eliminating bias and other sources of invalidity, on the one hand, and the insistence that interviews must be treated as situated interactions where the sense of respondents' talk is embedded in the context of the interaction.[67] This inconsistency will be explored more fully in section 4.3.6.

4.3.5 Qualitative interviewing compared to participant observation

There has been a long-standing and heated debate within qualitative research about the relative merits of qualitative interviewing and participant observation. Indeed, it has recently been claimed that qualitative researchers who are either unable or unwilling to engage in participant observation, and choose interviewing as an alternative, are marginalised by those who do (Kleinman *et al,* 1994). One of the earliest contributions to this debate was a paper by Becker and Geer (1960) in which they argued that participant observation should be regarded as a yardstick by which to measure all other research methods, as it represents the "most complete form of sociological datum" (322). They saw observation as overcoming some of the problems which they identified as inherent in the interview. These included:

- The problem of understanding what is said to the interviewer. Becker and Geer pointed out that, while interviewers and respondents may speak the same language, this does not mean that they can assume that they understand precisely what another person means by a particular word. There is a risk that, where researchers rely upon interview data, such misunderstandings are not uncovered.
- Difficulties related to respondents' unwillingness or inability to talk about certain matters. Becker and Geer argued that either deliberately or unawarely, interview respondents may fail to give interviewers all relevant information. They saw as a key advantage of participant observation its capacity for allowing the researcher to identify aspects of a setting which would otherwise be hidden from them.
- Problems arising from the way in which respondents are likely to see the situations

[65] See appendix 1 for details of this study; [66] The issue of validity in qualitative research is discussed more fully in chapter 5. [67] See section 3.2.1.3 for a discussion of context in qualitative research.

they report on "through distorted lenses", unlike the participant observer who can "check description against fact, and, noting discrepancies, become aware of systematic distortions made by the person under study" (328).

- Problems of inference. Becker and Geer argued that the process of inference is much greater in interviews than in participant observation. The interview researcher is obliged to make assumptions about the relationship of interview statements to actual events which may or may not be true.

Becker and Geer's position was challenged by Trow (1969) who argued that it was inappropriate to treat any method of data collection as inherently superior to others.[68] He held that, "the problem under investigation properly dictates the methods of investigation" (352). He took a surgical analogy to argue that scalpels are not necessarily better than forceps, unless you want to cut. In particular, he argued that participant observation is a relatively weak instrument for gathering data on sentiments, behaviours or relationships which are normatively proscribed.

Hammersley and Atkinson (1995) have pointed out that some of the criticisms that interactionists have directed at interviewing are based upon the assumptions of naturalism – that the best data are somehow untouched by human hands.[69] The problems raised by this appeal to naturalism are discussed above. As Silverman (1985) argued, "the opposition between artificial and naturally-occurring data is another methodological red-herring. Neither kind of data are [sic] intrinsically better than the other; everything depends upon the method of analysis" (156).

This debate remains far from settled. Denscombe (1983) identified a number of problems associated with exclusive reliance upon participant observation. He noted that, even where participant observation is the main data collection method, it may be important to interview respondents to check that the researcher's understanding of the situation corresponds with those of members.[70] Burgess (1985) argued that interviewing may be important in allowing the researcher to check inferences drawn from observations. Interviewing also offers the researcher the opportunity to collect data on groups that would be closed to him or her as a participant, such as those which require specific training or those from which (s)he is barred as a result of age, sex or race. Similarly, interviewing may allow the researcher to collect data on a wider range of settings than would be practical for observation and on events outside the immediate context which may have a direct bearing on that context. Hammersley and Atkinson (1995) commented that, "However skilful a researcher is in negotiating a role, some information will not be available at first hand" (125).

Kleinman and co-workers (1994) have argued that interviews have significant advantages over observational studies in certain circumstances. They suggested that interviews may be particularly useful when a phenomenon lacks a geographical base. While physically bounded phenomena (e.g. professional patient interactions in a clinic setting, see Silverman, 1981) may be amenable to participant observation, others (such as the child rearing practices of mothers of young babies in different ethnic groups, see Gantley *et al,* 1993[71]) may be less so. As McCracken (1988) pointed out, certain vital arenas of modern life, including the family, are unlikely to "suffer the presence of an observer for an extended period of time" (11).

Interviews are also seen by Kleinman and co-workers (1994) as allowing researchers to learn how members of a social group maintain, transform or challenge an identity in the face of certain experiences. For example, Charmaz (1983b) analysed qualitative interviews with 57 chronically ill adults to show that a narrow medicalised view of their physical discomfort failed to take account of the fundamental suffering associated with their loss of self as their former self-images crumbled away.

Interviews may also have the advantage of convenience, insofar as they can be scheduled to meet the constraints of both researcher and respondent (Denscombe, 1983). In situations where either the researcher or the respondent has limited time at their disposal, the prolonged contact required for participant observation may be impractical (McCracken, 1988). Crabtree and Miller (1991) suggested that qualitative interviews are a "rather useful technique for primary care research where time and money are frequent constraints and the topic can be fairly narrowly defined" (145).

On the other hand, Dingwall (1997a) has regretted the way in which the balance in qualitative research

[68] See section 3.1.1 for a discussion of the instrumental position on the choice between qualitative and quantitative methods; [69] See section 3.2.1.3; [70] See section 3.2.2.1; [71] See appendix 1 for details of this study.

has tipped from participant observation towards qualitative interviewing, as in an interview study: "We can pick and choose the messages that we hear and that we elicit. In observation, we have no choice but to listen to what the world is telling us" (64).

This preference for observation over interviewing is related to the critique of interview research (whether qualitative or quantitative), which is discussed below.

4.3.6 The radical critique of interviews

As discussed above, advocates of both standardised and alternative forms of interviews can be seen to be struggling with the same fundamental issue: 'How can we be sure that what respondents tell us represents reality?' While the answers offered by the two camps are clearly very different, the question is the same. Survey researchers emphasise the importance of controlling the context and content of the interview as rigidly as possible to maximise the chances that any differences recorded between respondents are real differences and not mere artefacts of the procedures used. Those who favour more qualitative approaches argue that the artificiality of standardised interview practice inhibits the expression of what the respondent really thinks and they call for interviews to be as much like everyday conversation as possible. In both cases, the aim is to discover what is **really** going on.[72] The validity of interviews is to be judged in terms of how carefully the interview has overcome the biases that are seen to be inherent in the situation, whether through standardisation or digging deep.

A more fundamental critique of the use of interviews to collect research data has emerged in recent years. Dingwall (1997a) traced this back to the 1960s and in particular to Cicourel's book *Method and Measurement in Sociology* in 1964 (Cicourel, 1964). Silverman (1985; 1993) has argued that this critique arises out of some of the inherent (though often unrecognised) strains within the very interactionist perspective which gave rise to the disillusionment with standardised interviews.[72]

This radical critique of interviews as a means of collecting data about external reality (whether biographical information, beliefs, ideas about what should be done, reports of present or past behaviour or conscious reasons for acting or thinking in particular ways, (Sellitz *et al,* 1964:265)) is seen by its proponents as the logical outcome of the understanding that interviews are essentially contextually situated social interactions. Drawing upon ethnomethodology, they have argued that **all** social interactions can best be seen as what Dingwall (1997a) described as "a dance of expectations". Dingwall outlined the basic features of the dance:

> I produce my actions in the expectation that you will understand them in a particular way. Your understanding reflects your expectations of what would be a proper action for me in these particular circumstances which, in turn, becomes the basis of your response which, itself, reflects your expectations of how I will respond. And so on. At any point, there may be disjuncture between actions, responses and expectations which requires that the parties engage in some sort of repair work. (56)

Within this perspective, social interactions are best understood as opportunities for **impression management** (Goffman, 1959), in which all parties strive to present themselves as competent and sane by those with whom they are interacting (Goffman, 1983 cited in Dingwall, 1997a).

When research interviews are treated as social interactions of this type the implications are radical. Rather than evaluating interview data as more or less accurate reports of external reality, we are obliged to view them as occasions when individuals feel called upon to give **accounts** of their actions, feelings, opinions etc., in such a way as to present themselves as competent, and indeed moral, members of particular communities. For example, the interview may be experienced as an occasion on which to display adequate parenthood, appropriate patienthood or competent professionalism. In addition, interviews are even more complex than naturally occurring social interactions insofar as respondents may feel additionally obliged to display competence as interview respondents.

The structure and content of everyday accounts is discussed by Scott and Lyman (1968), who define an account as "a linguistic device employed whenever an action is subjected to linguistic inquiry" (112). Clearly, this encompasses much of what occurs in the research interview, whether standardised or not. According to Scott and Lyman, accounts can be classified as either **excuses** or **justifications**, which arise when the possibility

[72] See section 3.2.1.3; [73] See section 4.3.2.

is raised that individuals have acted in some untoward manner. Accounts represent attempts to refute such challenges either by denying responsibility (excuses) or by arguing that the behaviour in question was understandable, given the situation (justifications).

Drew and Heritage (1992a) widened the application of the term 'accounts' to any occasion on which the "fabric of social order is put under stress" (Dingwall, 1997a). It is not necessary for individuals to have been directly challenged for them to feel obliged to produce account. Dingwall argued that it is an inescapable feature of interviews that the accepted social order is deliberately put under stress, as respondents are "required to demonstrate their competence in the role in which the interview casts them". This means that what is said in interview situations must be treated as an account, which represents the respondent's attempt to present self as a competent member of the community, rather than as literal descriptions of the respondent's reality.

The force of this perspective on interview data as accounts grounded in the context of their production is well illustrated by Stimson and Webb's classic study of consultations in British General Practice (Stimson and Webb, 1975[74]). Stimson and Webb combined participant observations in general practice consultations with interviews with patients. For practical reasons, they were unable to interview the patients whom they had actually observed interacting with doctors in the consultations. Nevertheless, they were struck by the inconsistencies between the reports that patients made about their interactions with doctors and those which they actually observed taking place. In particular, within the consultations, patients were generally passive and apparently compliant. They reported: "In general, the patient tends to be the more passive of the two ... Patients rarely give open expression to feelings of disagreement or dissatisfaction to the doctor's face but may mask them behind muttered or mumbled comments which are barely audible" (53).

By contrast, they found that atrocity stories were central to the accounts which patients gave of their interactions with doctors, in the context of the interviews. Such atrocity stories were dramatic presentations in which the patient was cast as hero and the doctor as incompetent. Contrary to Stimson and Webb's observations, these interview accounts presented the patient as active and the doctor as passive. In contrast to conventional approaches to the validity of interview data, which would tend to discredit the interview data as biased, Stimson and Webb analysed their data in terms of what the respondents could be seen to be **doing** in giving such accounts. They argued that atrocity stories offer patients a "vehicle for making the patient appear rational and sensible and for redressing imbalance between patient and doctor" (97).

As Silverman argued, Stimson and Webb's findings throw considerable doubt on the ways in which interview data have conventionally been used in healthcare research. For example, Waitzkin (1979) used interview data to argue that patients desired more information in medical encounters. Silverman pointed out that, in the light of Stimson and Webb's analysis, any such attempt to use interview accounts to evaluate non-interview settings is naive. Such interview data cannot be taken as evidence that patients evaluate medical consultations poorly, because patients, when interviewed, are likely to tell stories which cast them in a favourable light and such stories have an unknown relationship with what they expected or actually did in the situations they describe.

While this critique of the use of interviews as a source of data on external realities is now widely recognised, there is some difference of emphasis in terms of the implications which have been drawn about the use of interviews in social research. The most radical position is adopted by ethnomethodologists who argue that interviews can never be treated as a source of data for analysis of anything other than the interview itself. As Silverman (1985) observed some ethnomethodologists have been interested in studying the conversational sequencing of interviews and comparing this to natural conversation. However, they have rejected the possibility of treating interview reports as data on realities beyond the interview situation. In their own terms, ethnomethodologists may treat interviews as a topic of investigation but never as a resource (Garfinkel, 1967). In terms of this review, this position would imply that while interviews can contribute to our understanding of interview sequencing, their use as resources to illuminate people's use or opinion of various health technologies could never be warranted.

Silverman (1985; 1993), on the other hand, argued that, while interview talk cannot be treated as literal

[74] See appendix 1 for details of this study.

description of respondents' realities, they can nevertheless be treated as a resource as well as a topic: "When interviews take place, we witness both artful and possibly **universal** conversational practices **and** the display of cultural particulars expressing variable social practices" (Silverman, 1985:170 original emphasis).

In other words, by analysing what people are **doing** in their interview talk, rather than treating interviews as "passive filters towards some truths about people" (Baker, 1984), we can identify what the respondent takes to be self-evident about the world of which they are speaking. In this sense, interview data can be analysed in terms of the assumptions and the moral and cultural forms that they display. Silverman (1985; 1993) proposed, citing Voysey (1975)[75] and Baruch (1981),[76] that:

> ... research interviews offer access to a set of 'moral' realities firmly located in the cultural world. Once we rid ourselves of the palpably false assumption that interview statements can stand in any simple correspondence to the real world, we can bring fruitful analysis of the real forms of representation through which they are structured. (16)

A similar position is taken up by Dingwall (1981) who argued that the sociologist's job is not to "adjudicate between competing versions" but to understand the situated work which they do.

Silverman (1993) contrasted two studies of children with serious diseases to illustrate the difference that treating interviews as the occasions for displays of moral realities makes to the **analysis** of such data. He compares Burton's study of parents of children with cystic fibrosis (Burton, 1975) with Baruch's study of parents of children with congenital heart disease (Baruch, 1981[77]). As he noted, there are striking similarities between the accounts which the parents in both studies offered to the interviewers. In both studies, parents reported that, before the children were diagnosed, they had expressed concerns to health professionals about their babies, but that their concerns had been dismissed as groundless. However, the two researchers opted for very different approaches to the analysis of their data. Burton adopted an essentially interactionist perspective in which she treated the reports as at least potentially accurate reports of external events. By contrast, Baruch treated the interviews as local accomplishments and analysed them in terms of what the parents could be seen to be **doing** in the interviews. He argued that in the interviews the parents were displaying their moral responsibility. Hence, like Stimson and Webb, Baruch was concerned with the functions of the accounts that parents give, rather than treating them as straightforward reports on what had happened.[78]

Treating interview data as displays of respondents' perspectives, rather than as potentially accurate reports of external realities, has been shown to have considerable potential in informing health service provision. For example, Baruch (1981)[79] reported how he used his analysis of parents' accounts to demonstrate to the health professionals responsible for the paediatric cardiology clinic that the diagnosis of their child appeared to pose considerable moral and cultural problems for parents. On the basis of this analysis, an additional clinic was introduced, shortly after the diagnosis, for the express purpose of allowing parents to ask questions and "engage in a display of parental responsibility" (Silverman, 1985:175).

4.3.7 The interactionist response

Not all methodologists have accepted the force of this radical critique. For example, Hammersley and Atkinson (1995) have welcomed the increased emphasis on placing interview data within the context of their production and treating them as evidence of the perspectives of those who produce them. They accepted that accounts cannot be treated as "simply representations of the world; they are part of the world they describe and are thus shaped by the contexts in which they occur" (126). However, they rejected the idea that this means that interview accounts may never be read for what they tell us about the phenomena to which they refer. They insisted that people who participate in any setting acquire important knowledge about that setting which may be an important resource for the researcher. The validity of such data should not be taken at face value, any more than data from any source should be. Indeed, they argued that understanding the context in which accounts are produced, what people are doing in their accounting: "the presuppositions on which it relies, who produced it, for whom and why" (126), allows the researcher to anticipate potential biases that may threaten the validity of the information which such accounts contain. Hammersley and Atkinson (1995) concluded: "Separating the

[75] See appendix 1 for details of this study; [76] See appendix 1 for details of this study; [77] See appendix 1 for details of this study; [78] See section 3.2.2.1; [79] See appendix 1 for details of this study.

question of the truth or falsity of people's beliefs as social phenomena allows us to treat participants' knowledge as both resource and topic, and to do so in a principled way" (126).

The discontinuities between this and the more radical position outlined above should not be exaggerated. Hammersley and Atkinson argued that reports should not be taken at face value but should be assessed in the context of their production. Dingwall (1997a) argued that we can have no certainty that interview data represent literal descriptions of the respondents' reality. However, he accepted that:

> It may also be that some aspects of the respondent's reality can be glimpsed through the accounts: the selection of details, the choice of 'facts' in the narrative, perhaps. In real life we recognize that the accounts we receive every day contain some mix of the real and the representation and there seems no good reason why the accounts we receive as sociologists should be essentially different. My point is merely that interview data are fraught with problems because of the activity of the interviewer in producing them. (60)

Fundamentally, Dingwall, on the one hand, and Hammersley and Atkinson, on the other, agree that accounts are always grounded in the circumstances of their production. The difference lies in the implications which are drawn from this observation for the analysis of interview data. For Hammersley and Atkinson, these problems urge caution in the analysis of such data, but do not require us to abandon their use altogether. Likewise, Melia (1997), while recognising the limitations of interview data, nevertheless concluded that:

> If we can collect data with which to tell a plausible story, perhaps we should settle for that ... if the upshot [of phenomenological and postmodernist challenges] is methodological paralysis it might be better to take a more anarchic, or at least pragmatic approach ... with all that in mind, going off to interview people and coming back to tell Strong's 'plausible story' probably is, as he said, 'the best we can hope for'. (35–6)[80]

For Dingwall, the problems with interviews are so great as to mean that observation must always be treated as the method of choice.

4.3.8 Summary

In this section, we have presented the qualitative critique of structured interview techniques and we have discussed some of the advantages which advocates of qualitative interviewing have proposed for their methods. Such advantages include the claims that qualitative interviews are more successful in accessing respondents' interpretations, in penetrating public fronts and offer a more flexible research tool. We have also presented some of the debates within qualitative research concerning the relative merits of participant observation and interviewing, and the status of interview accounts. We conclude that interviews, whether standardised or non-standardised, cannot be treated as giving unproblematic access to respondents' perspectives and must always be analysed in relation to the circumstances of their production.

4.4 Documents[81]

In this section of the report, we focus upon the analysis of documents in qualitative research. As a number of authors have pointed out (e.g. Atkinson and Coffey, 1997; MacDonald and Tipton, 1993; Silverman, 1985; Hammersley and Atkinson, 1995; Douglas, 1967; Rees, 1981), textual sources have been relatively neglected within qualitative research, particularly within the British/American tradition. Atkinson and Coffey (1997) criticised much contemporary social research for its failure to take seriously the extent to which many of the settings and cultures under study are in fact "self-documenting" (45). The result, they argued, is that "occupational, professional, organizational and even academic settings are implicitly represented as devoid of written documents and other forms of recording" (45).

Various reasons have been given for the neglect of documentary sources in contemporary social research. First, Silverman (1985) suggested that the empiricism, which is characteristic of Anglo-Saxon research, is suspicious of the ephemerality and insubstantial nature of words and prefers to concentrate upon actions and the structures in which they are embedded (148). Second, authors such as Hammersley and Atkinson (1995) have suggested that the neglect of documentary sources in qualitative research can be traced to the focus on non-literate cultures in early social anthropology and upon the predominantly oral settings of "street culture and *demi-monde* beloved of many sociological field workers" (156). As these authors argued, the members of many of the settings

[80] This is a reference to Strong (1979a). See appendix 1 for details of this study; [81] See section 2.4.3 for an historical perspective on the use of documents in qualitative research.

that were studied by Chicago[82] sociologists (such as hobos, prostitutes and drug users) made no attempt to keep documentary records of their activities and made little use of written documents in their everyday lives. This led to the relative neglect of documentary sources of data. Third, Finnegan (1996) has suggested that social scientists have been reluctant to make use of documents because this source is shared with a number of other disciplines and therefore may not seem quite so distinctive as surveys, questionnaires and experiments.

It would be misleading to suggest that documentary analysis is completely absent from the history of qualitative research or indeed social research more generally. MacDonald and Tipton (1993) pointed out that the founders of the discipline of sociology (Marx, Durkheim and Weber) all made use of documentary analyses. Some of the classic studies that emanated from the Chicago School (e.g. Thomas, 1967; Thomas and Znaniecki, 1927) relied heavily upon the analysis of written documents such as letters and life histories.[83]

Whatever the historical or cultural reasons for the neglect of documentary sources, it is clear that documents are a major feature of contemporary society and, as such, an important source of data (Atkinson and Coffey, 1997). Finnegan and Thomas (1993) argued that:

> Both the proliferation of written records and communicating through writing more generally are widely seen as major features of modern society. Some scholars would go so far as to regard them as **the** defining attributes of Western industrial cultures, whether because of the (arguably) central role of print in our modern consciousness or through the development of modern bureaucracy with its reliance on written rules and administrative records ... it is certainly a fact that writing is one dominant medium in our culture for the storage, dissemination and retrieval of information. (98 original emphasis)

As Atkinson and Coffey (1997) pointed out, all modern organisations are highly dependent upon paperwork. People, at all levels of such organisations, are routinely involved in the production and consumption of written records. Any analysis of such organisations must take into account not only the contents of their records, but also the role of recording and retrieving such information within the organisation.

What is true of society in general is particularly true of healthcare settings, including hospitals, health authorities, general practices and other healthcare facilities, in which written documentation is a routine and pervasive activity. The analysis of such written records for research purposes has an important contribution to make to understanding healthcare settings. As Atkinson and Coffey (1997) argued:

> It is necessary to redress the balance [to include documentary analysis] if only for the sake of completeness and fidelity to the settings of social research. ... many organizations and settings have ways of representing themselves collectively to themselves and to others. It is, therefore, imperative, that our understanding of contemporary society ... incorporates those processes and products of self-description. (45)

4.4.1 Types of documentary sources

Various schemes have been suggested for categorising the types of documentary materials that are available for analysis (e.g. Denzin, 1970; Burgess, 1984; Finnegan, 1996). Hammersley and Atkinson (1995) suggested that such sources can be ranged along a continuum from **informal** to **formal** or **official**. At the informal end of this continuum there are the everyday accounts, which members of a setting produce in the course of their daily lives. These would include diaries, autobiographical accounts and letters. Even where, as if often the case, such documents are not created by the particular individuals who are the object of a research study, Hammersley and Atkinson argued that they can act as valuable stimuli for the researcher's imagination, as a source of hypotheses and foreshadowed problems. In recent years, there has been a growing interest in using such autobiographical materials for research purposes, particularly among feminist researchers, who have seen this as one way of re-establishing the link between the personal and the political (Stanley, 1992).

Hammersley and Atkinson (1995) suggested that informal documentary materials may be of particular use in suggesting comparisons, which will stimulate the insightful analysis of data: "There is every reason for the sociologist interested in, say, hospitals or clinics to examine works on a variety of other institutional settings – schools, courts, welfare agencies, religious houses, police stations, university departments, or emergency services, for example" (162).

[82] See section 2.4.3; [83] See section 2.4.3.

As it is rarely feasible for researchers to carry out empirical work in all of the settings that might offer fruitful comparisons with the setting under direct study, the use of documentary materials may be particularly useful here. As Hammersley and Atkinson (1995) acknowledged, the prime example of the use of such comparative methods is the work of Erving Goffman, whose analysis of mental hospitals was informed and enriched through comparison with other total institutions, such as army camps, monasteries, prisons and boarding schools.

As well as making use of the naturally-occurring documents that are relevant to a particular study, the researcher may also choose to generate documents, such as diaries and other types of written accounts, specifically for the purposes of research. Such accounts can offer the researcher the opportunity to study aspects of social life that might otherwise be inaccessible. Studies in which diaries have been used to collect such data include Coxon's study of sexual practices among gay males (Coxon, 1988), Davies and Atkinson's study of student midwives (Davies and Atkinson, 1991) and Robinson's study of illness behaviour among South Wales families (Robinson, 1971).

While some researchers have argued strongly for the analysis of documents on the grounds that failure to do so would involve neglect of an significant aspects of the settings under study, others (e.g. Burgess, 1984) have argued that the strength of documentary data lies in their capacity for giving access to the "individual's subjective view of social life"(Burgess, 1984:125). This approach to documentary analysis draws on the argument that qualitative research is particularly suited to uncovering members' meanings, which was critically discussed in section 3.2.2.

Psathas and co-workers (1968) elicited such documentary data from student nurses, as part of their study of students' impressions and attitudes about the role of the nurse. They showed two subgroups of student nurses (one at an early stage of their training and the other towards the end) a series of six slides depicting a student nurse in typical hospital settings. The students were asked to respond to the slides by writing a short story incorporating the slides. The researchers then compared the content of the stories produced by the two groups of students in relation to such features as the roles attributed to the various people depicted in the slides and the thoughts and feelings attributed to the student nurses. They concluded that, whereas the first-year nurses consistently explained what was happening by reference to how the patient was feeling, more advanced students adopted a more task-oriented approach, detailing the work that needed to be done for the patient. Psathas and co-workers recommended the elicitation of documents in this way as a means of overcoming students' reluctance and/or inability to articulate their own feelings as would be required in an interview situation.

Another example of a study that involved the elicitation of documents in healthcare settings is Bloor's study of death certification processes. This study (Bloor, 1994)[84] built upon an existing body of research which shows that there is considerable variation in the cause of death recorded upon death certificates. Bloor combined interviews with a sample of doctors responsible for a higher than average number of death certificates with the analysis of dummy death certificates, completed by the same doctors. These certificates were filled in by participating doctors, on the basis of standardised, short summaries describing the circumstances of death. Bloor found that doctors did differ in their decision about the cause of death in spite of the fact that they used the same case study information to arrive at their decisions. By combining interview data with the dummy death certification data, Bloor argued that he was able to gain insight into some of the reasons why death certification practices vary.

As well as such informal documents, qualitative researchers may also include more formal or official documents and administrative records in their analyses.[85] These would include public records such as compilations of official data and statistics and locally produced documents such as codes of practice, rotas, forms and other organisational materials. Such written documents are a pervasive feature of healthcare settings and yet relatively little attention has been given to them by researchers. As Hammersley and Atkinson (1995) suggested, many studies of medical settings treat such settings as if they were primarily oral cultures. Much research focuses upon professional–patient interactions but gives little attention to the activities of reading and writing, which are such pervasive features of the settings. They cited Rees (1981):

[84] See appendix 1 for details of this study; [85] Such records will, of course, include not only written materials but also those stored using electronic media.

> Both medicine and medical sociology have to a large extent neglected the record. Indeed, so rarely is it mentioned that one could be forgiven for thinking that medicine is a purely oral discipline. (55)

There are exceptions to this neglect of the official documentation and record. For example, Zerubavel drew on a range of documentary sources including timetables, work rosters and clinical rotations in his analysis of time in hospitals (Zerubavel, 1979, discussed in Hammersley and Atkinson, 1995). Similarly, Pettinari (1988) carried out a detailed analysis of the production of reports that surgeons write on operations, and the way in which junior surgeons learn to represent operations in ways that are deemed appropriate by their professional colleagues.

MacIntyre (1977) analysed the ways in which case sheet data were collected and recorded in the main antenatal clinic for a city with a population of about 180,000. MacIntyre noted the importance which obstetricians and gynaecologists attached to the information recorded upon patients' case sheets. Each case sheet was the joint production of five different members of staff (two nurses, a health visitor, an obstetrician and a booking clerk). MacIntyre observed the interviews carried out by these staff members and found that different members of staff elicited different and sometimes contradictory case sheet information from the same women. The case sheet data collected by different members of staff were found to relate to the occupational role of each and to their understanding of the usefulness of the information they were required to collect.

Barrett (1988) also reported a study of the process and consequences of document production in a healthcare setting. This research took place over a period of two years in a specialist unit for people diagnosed with schizophrenia, in Australia. He noted that the clinical records written on each patient by members of the psychiatric team played an important role within the work of the unit. He studied, in detail, the patients' admission assessments and the way in which the patients' information was entered into a clinical record, which then had a major influence upon the patients' future treatment. In particular, he noted the process by which lengthy, detailed verbal accounts, given by patients and their relatives, were reduced to sparse entries in the written record. He observed that this process transformed a detailed description into a textbook view of the signs and symptoms of schizophrenia.

Barrett's study also illustrated another approach to the analysis of documentary materials in healthcare settings, which involves focussing upon the use to which documents are put. He reported that the size of the patient's file effects how (s)he was perceived and responded to by staff. Patients in the unit were judged by the sheer bulk of their file, which was viewed by psychiatric staff as an indication of the chronicity of the patient's condition.

A further example of the ways in which documents are used in healthcare settings is found in the research carried out by Heath and Luff (1996). The starting point for this research was the observation that, in a variety of work settings, paper forms of information and documentation continue to play an important role, despite the introduction of computer technologies for recording, storing and retrieving information. Heath and Luff found that, in general medical practices, many GPs continued to use traditional, hand-written medical cards alongside new computer systems. They carried out an analysis of the information recorded on the hand-written records and the way in which GPs made use of these records in subsequent consultations. They found that, in spite of their brevity, record entries were an important resource in subsequent consultations. For the doctor, the order in which information was written, abbreviations and the inclusion or omission of certain items all provided information in a form that allowed the GP to both refer to it and add to it, without appearing to turn their attention away from the patient. At a glance, the doctor was able to gain a full sense of past consultations and any previous treatment.

Heath and Luff found that the inclusion of prescribing information was a particularly important interpretive resource for the doctor. Details of drugs prescribed were a vital indicator of the way in which the patients' symptoms had been interpreted at a previous consultation. The absence of any prescription details suggested that no physical cause for the complaint had been found. Heath and Luff observed that the importance of such prescribing information had been overlooked in the design of the computer software that was commonly used in general practice. In this system, diagnostic and prescribing details were stored separately and could not be accessed simultaneously. This meant that a vital aspect of the layout of the traditional written record was lost. Heath and Luff's analysis pointed to the limitations of current computer technology, which did not take account of the particular ways in which doctors make use of records within the consultation.

The third kind of documents that are available for analysis are visual documents, such as photographs,

advertisements and posters. In healthcare settings, researchers might choose to treat such artifacts as the posters in patient waiting rooms, health education materials and patient information leaflets as sources of data. As Finnegan (1996) suggested, while written texts may still be the primary form of documentary material, other sources, including radio and film material, maps, charts, photographs and audio and video sources are increasingly being seen as source material for documentary analysis.

One example of such use of visual documents in the health field is Stimson's analysis of drug advertisements (Stimson, 1977). Stimson focused upon drug advertisements directed at GPs in the UK. Statutory controls require that drug advertisements should provide doctors with sufficient information to reach an informed decision about whether, and under what circumstances, to use a particular drug. Stimson's aim was to assess the usefulness of advertisements as a source of therapeutic information for doctors. He carried out a preliminary content analysis, quantifying the amount of prescribing information contained in a sample of advertisements. This was followed up with a qualitative analysis of the same sample of advertisements, in which Stimson studied the visual imagery used and the ways in which such imagery conveyed a message about the drug in question. He found, for example, marked contrasts in the imagery used in advertisements for the contraceptive pill, on the one hand, and tranquillisers and antidepressants, on the other. In advertisements for the contraceptive pill, women were the main focus of the image and there was little or no background. In advertisements for tranquillisers and antidepressants, women were shown against a clear background, often showing a messy kitchen or a busy supermarket. These images were used to indicate the stress which might, it was implied, be relieved by the drug in question. Stimson's analysis suggests that, while drug advertisements may communicate relatively little therapeutic information, the visual imagery may be expected to have a significant impact.

4.4.2 Different traditions in documentary analysis

Jupp and Norris (1993) identified three approaches to the analysis of documentary data: positivist, quantitative content analysis; the interpretative tradition; and the critical tradition.

4.4.2.1 Positivist, quantitative content analysis

Jupp and Norris drew upon Lasswell and co-workers' formulation of content analysis as, "who says what, to whom, how and with what effect?" (Lasswell *et al,* 1952:12 in Jupp and Norris, 1993:38). They argued that this approach to documentary analysis dominated the field until the 1960s. Here documents were treated as an objective means for making inferences about the realities which they represent. The methodological principles underlying content analysis were itemised by Holsti (1969:5 cited in Jupp and Norris, 1993:40). The procedures should be objective (i.e. carried out in terms of explicitly formulated rules) and systematic (rules must be applied with consistency). Content analysis must have theoretical relevance. Content analysis is typically quantitative and is concerned with the manifest and surface meaning rather than making any attempt to uncover deeper levels of meaning. As Jupp and Norris suggested, a distinctive feature of such content analyses is their unquestioning stance towards the contents of the documents being analysed and a lack of concern with the role of documents as a means of maintaining existing power relations. As we shall see, these features mark them out from the other two traditions of documentary analysis.

4.4.2.2 The interpretative tradition

Jupp and Norris presented the interpretative tradition in documentary analysis as arising out of a critique of the representational model of content analysis. Rather than being treated as a representation of reality, within the interpretative tradition documents are more likely to be seen as significant social mechanisms, which can be analysed as realities in their own right rather than as windows on the reality they purport to represent. The implications of this position are more fully discussed in section 4.4.3.

4.4.2.3 The critical tradition[86]

While the methods adopted by critical theorists may not differ greatly from those adopted by interpretivist, the focus of their research is substantially different. The emphasis is upon the role of official documents in conflicts between different social groupings and the way in which documents and texts can be used by one group to regulate the actions of another. Documents are studied as legitimating devices and critical theorists seek to analyse them in terms of their relation to ideology, power and control.

Discourse analysis represents an important approach to the analysis of documentary material within the critical paradigm. Central to discourse analysis is the

[86] See section 3.1.3 for a discussion of critical theory.

argument that "knowledge and discourses must be analysed in terms of different points or stages in history, and also in terms of social relations prevailing at these points rather than in terms of individual constructions" (Jupp and Norris, 1993:47). Discourses are frequently in conflict with one another and some discourses are more powerful than others. There is a strong link between the production of discourses and the exercise of power within society. Dominant discourses give shape to what is seen as right or wrong in society and shape the solutions which are preferred.

4.4.3 The status of documentary data

The status given to documentary evidence varies across different qualitative research traditions. Finnegan (1996) distinguished between direct and indirect uses of documentary material. Direct use involves treating the document as an at least potentially accurate record of some phenomenon. For example, the researcher might use a hospital's personnel department records to obtain data on staffing levels in various hospital departments. Here the researcher is seeking to use records to reconstruct events that occurred during times when or in places where the researcher was not present, but which may be relevant to the study (Burgess, 1984). For researchers taking this position, methodological concerns revolve around the trustworthiness of the document as an accurate representation of reality. Burgess (1984) suggested that the researcher who wishes to use documents must be concerned about the authenticity of such documents, with particular attention to the risks of forgery or mis-representation. In addition, the researcher should be aware of the risks of distortion and deception. For Burgess, the central question appears to be, 'How can we know that the document is not the product of an agent who is attempting to distort or deceive?' Burgess recommended a number of strategies which the researcher may use to identify documents which are suspect in this way.

Finnegan (1996) discussed the indirect use of documents as an alternative approach. Here the researcher is not so much concerned with the truth or falsity of the documents' content, as with treating the document as an account that can be analysed for what it tells us about the perspectives of the members of particular settings and the contexts in which they operate. Just as, in the section of this report dealing with interview accounts,[87] we discussed the argument that interviews should always be understood as the product of a situated encounter between the interviewer and the respondent, so too documents can be analysed as the product of the context in which they were generated. As such, they may tell us more about the context than they do about the substantive issues which they seek to report. Finnegan (1996) put it this way:

> The **gloss** put on the message can itself convey indirect information about, say, the ideals aimed at, the standard terminology used in a particular place or period, the kinds of subterfuges engaged in, or the sort of images though likely to appeal to the intended market. (143 original emphasis)

Atkinson and Coffey (1997) argued that documents cannot be treated as "transparent representations of organizational routines, decision-making processes or professional diagnoses". Such documents, they argued, should not be used as 'surrogates' of other kinds of data. Documentary evidence cannot be treated as a representation of how an organisation actually works. However, this does not mean that such documents are unimportant. Rather, Atkinson and Coffey suggested:

> We have to approach them [documents] for what they are and what they are used to accomplish. We should examine their place in organizational settings, the cultural values attached to them and what they are used to accomplish. The analysis of such evidence is therefore an important part of many ethnographic studies of organizations, professional work and similar settings. (47)

Again, Hammersley and Atkinson (1995) suggested, just because documentary accounts cannot be treated as accurate or objective does not mean that they cannot be analysed in terms of what "they reveal about the teller's interests, perspectives and presuppositions" (160). For example, Miller and Morgan (1993) have analysed *curriculum vitae* as "presentations of self" (Goffman, 1959), which tell us as much about the circumstances under which such documents were produced as about the individuals whom they purport to describe. In this way, documents can be treated as self-presentations whose analysis must incorporate "a clear understanding of how documents are produced, circulated, read, stored and used for a wide variety of purposes" (Atkinson and Coffey, 1997). Atkinson and Coffey recommended that, in analysing documents, the researcher should pay attention to the form and function of the texts under analysis, rather than the accuracy or otherwise of their content. In this way such documents could be analysed in terms of their rhetorical features, the

[87] See section 4.3.6 for a discussion of accounts.

process through which they were produced and their intended readership.

4.4.4 Summary

Documents are relatively neglected source of research data. This is a particular problem in settings such as hospitals and other treatment facilities, where written documentation plays a central role in the organisation of activities within the settings. In this section we have discussed the range of documentary materials that are available to researchers, and the alternative theoretical traditions which underpin different approaches to documentary analysis. We have also discussed an important distinction between those researchers who set out to treat documents as more or less accurate representations of reality, and those who seek to analyse documentary materials in terms of their form and the functions which they fulfil in the contexts in which they are situated.

4.5 Conversation analysis

The past 25 years have witnessed the emergence of a growing body of qualitative research involving the analysis of audio and audio visual recordings of naturally occurring social interaction. For the analysis of social interaction, these data offer a number of advantages over more traditional modes of inquiry such as field observation, interviews and questionnaires. Heath and Luff (1992a) described these as follows:

> Recordings of human interaction provide the researcher with access to the richness and complexity of social action, allowing particular events to be scrutinised repeatedly and subjected to detailed inspection. They provide raw data to which a range of analytic interests can be applied, unconstrained by the concerns of a particular research project. Moreover, audio and video recordings enable other researchers within the scientific community to evaluate the strength of particular analyses with respect to the raw data and thus provide an important constraint on the quality and rigour of findings and explanations. Audio and video recordings give researchers a cheap and reliable technology which provides **repeatable** access to specific details of real world actions, activities and events; a microscope with which to study human life. (307–8 original emphasis)

Whilst studies of audio and audio-visual recordings of social interaction have been conducted from a variety of perspectives, there has been a rapid growth in the use of CA. CA emerged in the 1960s, as part of the research programme of Ethnomethodology[88] which developed from a series of seminal studies conducted by Harold Garfinkel (1967). In contrast to interactionist studies, ethnomethodological studies are not primarily concerned with the subject's point of view.[89] Instead they explicate the routine, often tacit, practices that participants use in accomplishing everyday activities. They focus on how people do things, rather than on how people see things (Maynard 1989).

Some ethnomethodological researchers seek to identify participants' practices through traditional observational methods. However, the pioneering research of Sacks and his colleagues, Schegloff and Jefferson, has led to the emergence of a substantial corpus of ethnomethodological studies concerned with the analysis of audio and audio-visual recordings of social interaction. These studies were inspired, in particular, by Sacks's proposal that the analysis of audio recordings of talk-in-interaction provides the possibility of developing a "naturalistic observation discipline which (can) deal with the details of social action(s) rigorously, empirically, and formally" (Schegloff and Sacks, 1973:233). The methodology developed in these studies came to be known as CA. Although CA was initially pioneered within sociology, its approach and findings are now used by scholars in a range of disciplines, including anthropology, Computer Supported Cooperative Work, Human Computer Interaction, Social Psychology and Interactional Linguistics.

CA research describes the interactional practices that speakers use and rely upon in producing their own conduct and interpreting and dealing with the conduct of others. In the words of Goodwin and Heritage (1990): "[it] seeks to describe the underlying social organization – conceived as an institutionalized substratum of interactional rules, procedures, and conventions – through which orderly and intelligible social interaction is made possible" (283).

Thus conversation analysts are interested in how speakers open and close their interactions, allocate opportunities to speak, introduce and change topics, manage disagreements, deal with problems in speaking, hearing or understanding, and the like. They also investigate the social organisation of visual conduct, such as gaze and gesture,

[88] See section 2.4.4; [89] See section 3.2.2.1.

paying particular attention to the ways in which it is related to spoken interaction (e.g. Goodwin, 1981; Heath, 1986).

CA research does not involve the formulation and empirical testing of *a priori* hypotheses. Instead it identifies patterns of verbal and/or non-verbal interaction through the use of the procedures of analytic induction.[90] Heritage (1997) described the mode of analysis as follows:

> As competent language users, analysts develop more or less conceptually informed hunches about the uses and organisational properties of particular conversational practices. At this point, the work normally begins with an inductive search for instances of the practice under investigation using as wide a range of data as possible... Once possessed of a set of cases that appear to embody a conversational practice or procedure, the detailed work of specifying the scope and limits of the practice begins. A major component of this involves 'deviant case analysis'. This involves examining cases where the general pattern is departed from and examining whether, and in what ways, the participants orient to such departures. Used in this way, deviant case analysis is an important resource for determining whether the basic pattern simply embodies an empirical regularity that happens to occur, or whether it involves something that is oriented to as a normative interactional procedure. (399)

In undertaking this form of "pattern and deviant case" analysis[91] (Heritage 1997:399), CA researchers repeatedly replay audio or video recordings of naturally occurring interaction, carefully transcribing the events. The transcripts capture not only what is said, but also a variety of details of speech production, such as overlapping talk, pauses within and between utterances, stress, pitch and volume. They may also track visual conduct such as gaze direction and gesture. These transcripts facilitate the fine-grained analysis of the recordings, enabling researchers to reveal and analyse tacit, 'seen but unnoticed' aspects of human conduct which would otherwise be unavailable for systematic study. Moreover, extracts from the transcripts are included in research reports as exemplars of interactional phenomena under investigation.[92]

It is important to note that CA researchers develop analyses that are sensitive to the orientations of the speakers themselves. They do this by anchoring their observations in the understandings that participants unavoidably display of each other's conduct during their interactions. Thus Sacks, Schegloff and Jefferson (1974) suggested the following:

> [It] is a systematic consequence of the turn taking organisation for conversation that it obliges its participants to display to each other, in a turn's talk, their understanding of the other turn's talk. More generally, a turn's talk will be heard as directed to a prior turn's talk, unless special techniques are used to locate some other talk to which it is directed...But while understandings of other turns' talk are displayed to co-participants, they are available as well to professional analysts, who are thereby afforded a proof criterion (and search procedure) for the analysis of what a turns' talk is occupied with. Since it is the parties' understandings of prior turns' talk that is relevant to their construction of next turns, it is their understandings that are wanted for analysis. The display of those understandings in the talk in subsequent turns affords a resource for the analysis of prior turns, and proof procedure for professional analyses of prior turns, resources intrinsic to the data themselves. (728–9)

Although CA began from the study of ordinary conversations, its approach and findings have been widely used to analyse "interaction in 'institutional settings where more or less official or formal task- or role-based activities are undertaken" (Heritage 1997:406).[93] Thus, for example, CA studies have focussed on interactions in medical consultations, broadcast interviews, telephone calls for emergency assistance, business meetings, classroom lessons, divorce mediation sessions, small claims courts, and psychiatric intake interviews (e.g. Atkinson and Heritage, 1984; Boden, 1994; Button and Lee, 1987; Boden and Zimmerman, 1991; Drew and Heritage, 1992). These studies examined how social institutions are evoked and managed in interaction. Specifically, they showed that participants made particular institutions and their associated identities relevant within their talk, and how the invocation of such institutions and identities constituted both a constraint on and a resource for the activities in which the participants were engaged. This generally involved analysts describing how participants adapted a limited number of the full range of generic speaking practices which were available to them to manage tasks and constraints that are indigenous to particular institutional settings.

[90] See section 4.6.1.3; [91] See section 5.3.7; [92] For introductions to conversation analysis, see Goodwin and Heritage (1990), Greatbatch *et al,* (1995b), Heath (1997), Heath and Luff (1993), Heritage (1988; 1995; 1997) and Zimmerman (1988); [93] Thus, despite its name, conversation analysis represents a generic approach to the study of social interaction.

The past 10 years have witnessed the emergence of video-based studies of naturalistic work and interaction which, whilst drawing on CA, adopt a broader focus than the studies of institutional talk mentioned above. Whereas the latter studies are primarily interested in the role of talk in the accomplishment of social activities, these studies also analyse the use of tools, texts and technologies. The objective is to explicate the informal, taken-for-granted practices and reasonings, which personnel rely upon in accomplishing and coordinating workplace activities. Topics have included the ways in which people coordinate apparently individual tasks with the actions of others, how personnel monitor each other's conduct for its relevance to their own work and how the use of various artifacts shapes and constrains interpersonal communication. Video-based studies, which have drawn from CA in order to investigate the social organisation of work, interaction and the use of new technology have been conducted in a broad range of settings, including offices (Suchman, 1987), airport operation rooms (Goodwin and Goodwin 1992), urban transport control rooms (Heath and Luff 1992a, 1997), City trading rooms (Heath *et al,* 1995) and emergency dispatch centres (Whalen 1995). This body of research has demonstrated that, although the approach of CA cannot be simply applied to activities in such settings, as those activities are rarely accomplished wholly through talk, it nonetheless provides the foundation for the explication of the social interactional organisation of work, interaction and technology (Greatbatch *et al,* 1995b).

In contrast to more traditional CA research, these video-based analyses of workplace activities often rely on information gathered through observation, and interviews. This has proved necessary because the patterns of interaction and the organisation of activities in complex workplace settings often cannot be discerned or understood solely from analysis of video recordings. However, the use of such ethnographic information remains highly contentious because it may be used to warrant claims that have not been grounded in the understandings that interactants themselves display to each other (Hopper, 1991; Maynard and Clayman 1991).

In sum, CA comprises a data-driven methodology (Clayman and Maynard, 1995), which seeks to ground theoretical and analytic claims in the orientations that subjects themselves display to one another during their interactions. CA researchers explicate the interactional practices through which people manage their interactions in an orderly, intelligible fashion. Some CA researchers also examine the relationships between these practices and wider social and institutional formations. Others adopt a still broader focus by considering how various artifacts feature in everyday social action and interaction. Together these strands of research demonstrate how the analysis of audio and audio-visual recordings of naturally occurring social activities reveals tacit "seen-but-unnoticed background features" (Garfinkel 1967) of social settings which remain largely, if not wholly, unavailable to researchers who rely solely on data generated through observation, interviews or questionnaires.

Chapter 6 contains a case study, which considers the application of interaction analysis to research on the introduction of medical information systems.

4.6 Qualitative data analysis

Qualitative research has the propensity for generating data which Bryman and Burgess (1994:216) have described as being "voluminous", "unstructured" and "unwieldy". Researchers can find themselves faced with a situation where their fieldwork produces vast quantities of information and can fall victim to the problems of "data overload" (Miles and Huberman, 1994:2). Dealing efficiently and productively with one's data is highly demanding. However, until relatively recently, the analysis of qualitative data attracted little methodological discussion and the details of qualitative data analysis practice also remained largely shrouded in mystery. This gave rise to two distinct problems. Researchers lacked points of reference to which they could turn to learn about methods of analysis, and doubts were cast upon the validity of findings where no method of analysis had been explicated.[94]

> ... the full weight of evidence for a given conclusion is not usually presented. The observer's conclusions often have a *prima facie* validity, a 'ring of truth' but the reader of his research report has no way of knowing whether a solid basis of fact underlies this. The reader does not have the data available with which to convince himself and must rely on his faith in the researcher's honesty and intelligence. (Becker and Geer, 1960:270 commenting on findings from participant observation studies.)

[94] See section 5.3.2.

> The most serious and central difficulty in the use of qualitative data is that methods of analysis are not well formulated. For quantitative data, there are clear conventions the researcher can use. But the analyst faced with a bank of qualitative data has very few guidelines for protection against self-delusion, let alone the presentation of unreliable or invalid conclusions to scientific or policy-making audiences. How can we be sure that an 'earthy' 'undeniable' 'serendipitous' finding is not, in fact, **wrong**? (Miles 1979 cited in Miles and Huberman, 1994:2 original emphasis.)

The last 15 years, however, have seen a move away from what Anselm Strauss (1987) graphically described as "the long era of flying by the seat of one's pants!" (Preface), and there are now texts which serve as useful handbooks for both students and researchers, (e.g. Strauss, 1987; Dey, 1993). There has also been a significant growth in writings which explore the theoretical and practical issues of qualitative data analysis.

In addition to providing an overview of the key issues, the following sections will also seek to emphasise the ways in which the process of qualitative data analysis is distinct from that employed within quantitative research. Before examining specific methods and elements of the process, we draw attention to three significant aspects of qualitative data analysis.

The first concerns confusions in terminology. Qualitative researchers may refer to the process of handling information after collection as **analysing** (Bryman and Burgess, 1994), **interpreting** (Silverman, 1993), **transforming** (Wolcott, 1994), or **making sense of** (Hammersley and Atkinson, 1995) their data. The term 'analysis' itself is also potentially confusing as it is used to indicate a range of different procedures. Tesch (1990) identified three types of qualitative analysis:

- analysis based on language (e.g. discourse analysis, symbolic interactionism, ethnomethodology)
- descriptive/interpretive analysis (e.g. classic ethnographies, life histories)
- theory building types of analysis (e.g. grounded theory).

However, it is important to note, as Bryman and Burgess (1994) have stated, that some of these types can overlap. For example, symbolic interactionists may also want to theory build from their data. Wolcott (1994) made a useful distinction between the use of the term 'analysis' as data **management**, indicating systems that allow data to be handled more efficiently, and as **procedure** in which features and relationships are identified. However, this review will consider both aspects together since they tend to occur simultaneously as research progresses.

Second, we identify three differences between qualitative and quantitative analysis.

- Both quantitative and qualitative research incorporate the same components of research design, sample selection, data collection and analysis. However, whether qualitative research is focussing on description and/or explanation, or on theory emergence and/or testing,[95] the data collection and analysis stages tend to be **iterative** rather than sequential (Glaser and Strauss, 1967; Schatzman and Strauss, 1973; Merriam, 1988; Marshall and Rossman, 1989; Fielding, 1993; Maykut and Morehouse, 1994; Hammersley and Atkinson, 1995). Miles and Huberman (1994) defined analysis as consisting of "three concurrent flows of activity: data reduction, data display and conclusion drawing and verification" (10). The findings of preliminary analyses of data from the first phases of fieldwork will serve to inform thinking on subsequent modifications of the research design, though analysis will continue after data collection has been completed.

  > The process of data collection and analysis is recursive and dynamic. But this is not to say that the analysis is finished when all the data have been collected. Quite the opposite. Analysis becomes more intensive once all the data are in, even though the analysis has been an ongoing activity. (Merriam, 1988:123.)

- Quantitative analysis methods seek to reduce data in order to make their meaning more accessible. For example, it is not necessary to read through all the individual replies collected in questionnaires from a large scale survey because the responses can be summarised by generating a series of statistics. Qualitative analysis also aims to make the data more manageable, but presents findings by expansion and development of ideas rather than by reduction. Analytic induction[96] and theoretical sampling,[97] both of which are discussed in detail elsewhere in this report, were cited by Ragin (1994) as examples of techniques which he terms **data enhancers**. He stated:

[95] See section 4.6.1; [96] See section 4.6.1.3; [97] See section 4.1.1.4.

> Most quantitative data techniques are data condensers. They condense data in order to reveal the big picture … qualitative methods, by contrast, are best understood as data enhancers. When qualitative methods are used to enhance social data, researchers see things about their subjects that they might miss otherwise. Data enhancement is the key to in-depth knowledge. Almost all of qualitative research seeks to construct representations based on in-depth, detailed knowledge of cases, often to correct representations or to offer new representations of the research subject. Thus qualitative researchers share an interest in procedures that clarify key aspects of research subjects – procedures that make it possible to see aspects of cases that might otherwise be missed. (92)

- In order for ideas to be expanded and developed, the qualitative data analysis process will always involve a strong element of **individuality** and this will stem from the uniqueness of both the researcher/s and the research setting. All those involved in any type of research come to a new project carrying a variety of individual experience and background knowledge. However, in qualitative work this is particularly significant. Glaser and Strauss (1967) described the way in which **theoretical sensitivity** is essential to the analysis process:

  > ...the sociologist should be sufficiently theoretically sensitive so that he can conceptualize and formulate a theory as it emerges from the data. Once started, theoretical sensitivity is forever in continual development. It is developed as over many years the sociologist thinks in theoretical terms about what he knows and as he queries many different theories ... theoretical sensitivity of a sociologist has two other characteristics. First, it involves his personal and temperamental bent. Second it involves the sociologist's ability to have theoretical insights into his area of research, combined with an ability to make something of his insights. (46)

Strauss and Corbin (1990) also described the importance of theoretical sensitivity which is derived, not only from knowledge of the literature on the subject under study, but also from both the professional and personal experience which a researcher brings to the analysis. Theoretical sensitivity, which increases as one interacts with the data, is automatically employed in decisions concerning the naming of categories and making connections between them. However, Strauss and Corbin also stressed the need to maintain a balance between individuality and science in order to produce theory that is both reliable and valid. In order to achieve this one must "hold on to the reality of a phenomenon, rather than just thinking imaginatively about it" (44). This is achieved through continuous reflection on the fit between data and reality, by regarding all theoretical explanations, categories, hypotheses and questions about the data as being provisional until they are checked out against the data, and by following research procedures correctly in order to sustain rigour. Strauss and Corbin stressed that grounded theory[98] is a scientific method because its procedures meet the necessary criteria for generalisation, verification and rigour, but that these must not be interpreted in positivistic terms because of the incorporated individuality of the researcher.

It is the elements of individuality and uniqueness that make a complete standardisation of method in qualitative analysis impossible. Miles and Huberman (1994) see only a minor risk of formalisation as qualitative practices are now increasingly "dissected and reassembled" because they consider this particular brand of analysis to be "more a craft than a slavish adherence to methodological rules" (5).

Third, there are considerable differences between the coding process in qualitative and quantitative research. The need for systematic methods of handling qualitative materials in order to avoid problems of overload has already been emphasised. Although not all qualitative research is concerned with the generation of theory,[99] it is necessary to label concrete categories and analytical ideas as they emerge from the data in studies where one wishes to facilitate theory development. Both of these activities are addressed through the activity of coding.

In quantitative analysis, coding categories reflect prior hypotheses and theoretical assumptions. In qualitative analysis, however, researchers organise their data and simultaneously begin to identify its relevant elements by cataloguing the substantive topics and attaching labels to concepts and themes which they see emerging from the material as they work with it. Such processes are usually referred to as indexing and/or coding of the data, but the use of these terms differs and the operation may involve either two or three stages.

For example, according to Miles and Huberman (1994), codes are basically "tags or labels for assigning units of meaning to the descriptive or inferential information compiled during a study"

[98] See section 4.6.1.4; [99] See section 4.6.1.2.

(56). They identified two stages of the coding process. First-level coding can either involve the use of a provisional list of codes (i.e. *a priori* codes), which have been drawn up before fieldwork has been carried out, or wholly inductive codes, which are developed from ideas that emerge from the initial collections of data, or a 'general accounting scheme, which is constructed at a midway point and is not content-specific but points to the general areas in which codes can subsequently be developed inductively. Pattern-level coding groups the summaries obtained in first-level coding into a smaller number of analytic units and facilitates the focussing of later fieldwork.

Charmaz (1983a) defined coding as "simply the process of categorizing and sorting the data" (111). Categories may range from those applied to simple concrete topics to more abstract conceptual classifications which are indicative of emerging theory. However, according to Charmaz, coding provides "the pivotal link between data collection and its conceptual rendering (and therefore) becomes the fundamental means of developing the analysis" (112). She also described a two-stage process. Initial coding is a summarising activity, where the researcher looks for what can be discovered in the data, and here the development of *in vivo* codes, which are constructed from the respondents' own language, may be important. Focused coding involves the analytical application of the coding framework to the larger data set and it is during this stage that theoretical sensitivity (see above) becomes increasingly important as the researcher continues to explore the emerging analysis and interpret the meaning of the data.

Richards and Richards (1994a) distinguished between coding for text retrieval and coding for theory generation. They presented coding for text retrieval as being essential in order to control the volumes of rich data, which qualitative research methods typically collect, but recognised there is a problem with this in that it can pre-empt the analysis because information becomes **boxed**. They stressed that, if one is aiming to generate theory rather than test it, one must remain open to new ideas, categories and concepts and making links between them. It was these ideas that led to the development of the NUD*IST[100] software package for qualitative data analysis.[101]

Strauss and Corbin (1990), however, defined coding as a practice which encompasses the whole process of analysing data – "coding represents the operations by which data are broken down, conceptualized and put back together. It is the central process by which theories are built from data" (57). These authors presented very detailed outline of a three stage coding process. During open coding, data are broken down, conceptualised and categorised, thus allowing the researcher to identify concepts and form categories by grouping together similar incidences and events. The data are then put back together in new ways whilst searching for additional properties of each category and making connections between them; this activity is termed **axial** coding. The researcher subsequently undertakes selective coding, which is "the process of picking out the core category, systematically relating it to other categories, validating those relationships and filling in categories which need further refinement and development" (116). The creation of the core category is essential for the development of a clear analytic story from the data because it represents the central phenomenon around which all other categories are integrated.

Approaching the issue of qualitative analysis with a different focus, Lonkila (1995) emphasised that in methods such as grounded theory,[102] coding refers to the whole process of analysing the data, whereas in discussions of computer-assisted qualitative data analysis software (CAQDAS), the term coding refers to the mechanical process of attaching keywords to text segments. The use of qualitative analysis software will be discussed after the following section which describes the practice of qualitative data analysis.

4.6.1 The practice of qualitative data analysis[103]

4.6.1.1 Description and explanation

Some types of qualitative research aim to produce description. For example, in ethnographic studies, data analysis will be focused towards the production of a detailed record of the research setting and of individuals' interpretations of their world. Others are concerned with theory generation and testing and therefore employ different methods of data analysis.

Strauss and Corbin (1990) outlined three approaches to analysis, which are arranged along a continuum in terms of the level of data interpretation:

[100] Non-numerical Unstructured Data Indexing, Searching and Theorising software package; [101] See section 4.6.2.1 for a discussion of computerised qualitative data analysis; [102] See section 4.6.1.4; [103] See section 2.1 for an historical perspective on the debate about the nature of causal explanation in social research.

- presentation of data without analysis where "the informants speak for themselves", thus aiming to give what is regarded to be an "honest account with little or no interpretation" (21)
- analysis which aims to provide an accurate description of the subject under study; this approach presents a "rich and believable" account where the commentary may also be of a theoretical nature (22)
- development of theory where conceptualisation of data and relation of concepts is undertaken in order to "form a theoretical rendition of a reality" (22).

Hammersley and Atkinson (1995) described the practice of progressive focussing which involves:

> a gradual shift from a concern with describing social events and processes towards developing and testing explanations or theories. However different studies vary considerably in the distance they travel along this road. Some remain heavily descriptive, ranging from narrative life histories of an individual group or organization, to accounts of the way of life to be found in particular settings. (207)

Hammersley and Atkinson stressed that such work involves selection and interpretation, but that theory remains implicit and is "used as a tool, rather than forming the focus of the research". Some researchers will move on from description and work towards the formulation of explanations or theoretical models by building up general and analytical categories based on features of the phenomenon under study which help to lead to a greater understanding of that which is being indicated by the data.

Descriptions and explanations from individual study findings can be used as the basis for further work. They may be pieced together in order to build up a composite picture, jigsaw puzzle or map, or in other words to provide a panoramic view of, for example, a whole city, system or society (Hammersley *et al,* 1985:51). It is also possible to develop a **taxonomy**, which will contain sub-sets of a general category, in order to present a way in which findings from a current study may be applied elsewhere. Hammersley and Atkinson cited the example of Glaser and Strauss' taxonomy of awareness contexts which was developed from their work with terminally ill hospital patients (Glaser and Strauss, 1965b). In this taxonomy **closed awareness** refers to situations where the patient is not informed of the diagnosis and prognosis and, at the other end of the continuum, **open awareness** to where both the patient and their families are fully informed (Hammersley and Atkinson, 1995:216).

4.6.1.2 Qualitative data analysis and the development of theory

In quantitative studies, where the **testing** of theory is emphasised, the hypothetico-deductive method of data analysis predominates.[104] This approach, which Strong (1979a)[105] termed the **big bang mode of analysis**, separates theory-development from theory-testing. Hypotheses and theories are formulated **before** fieldwork commences and generalisations and predictions are made on the basis of deductions from them. Successful prediction, as confirmed by empirical findings, is seen to be a necessary test of the adequacy of a hypothesis, or as Strong puts it "one systematically deduces a set of hypotheses from a body of theory and attempts to falsify them at one go" (234). Theory and explanation are seen to be achieved once successful predictions can be made because, if things are confirmed as expected by the data, then the original theory is upheld. If the data disproves the hypothesis, then it must be revised or rejected.

In qualitative research, it is much more common to find that the processes of theory development and empirical inquiry are interwoven. The "delineation of theoretical ideas is usually viewed as a phase that occurs during or at the end of fieldwork, rather than being a precursor to it" (Bryman, 1988:81). The **grounded theory** and **analytic induction** methods of analysis, which are discussed below,[106] both aim to generate theory from data. It must be emphasised that in studies that purport to use either of these techniques, there is often a gap between programmatic statements of method and actual practice. In many cases the practice may incorporate the **logic** of the approach whilst not necessarily including every stage of the process.

4.6.1.3 Analytic induction

Analytic induction has been described as a method:

> in which there is a systematic search for falsifying evidence and modification of the theory until no further disconfirming evidence can be found. This strategy may be employed both within a study of a

[104] See section 3.2.1.2 for a discussion of the roles of induction and deduction in qualitative research; [105] See appendix 1 for details of this study; [106] See sections 4.6.1.3 and 4.6.1.4.

> particular setting to check the validity of descriptions and explanations of processes there, or on a larger scale to try to establish universal patterns of causality across all relevant groups and settings... (Hammersley, for the DE304 Course Team, 1979a:28.)

In this section we will outline Znaniecki's original formulation of the method of analytic induction, together with the two classic examples of its implementation (Znaniecki, 1963). Critiques of Znaniecki's claims will then be examined and finally, examples of later studies, which have attempted to apply the logic of analytic induction, rather than to conform to the practice of the method in its entirety, will be discussed.

The process of analytic induction was originally described by Znaniecki who claimed it was the "true method of physical and biological sciences" (Hammersley and Atkinson, 1995:234) and that it was superior to the statistical method of analysis because of its capacity for producing universal laws that enable one to state both the necessary and **sufficient** conditions under which a phenomenon occurs. He challenged the suitability of using enumerative induction, or statistical methods, to study relationships in social science because these merely produce correlations to which there are always exceptions.

Analytic induction is concerned with presenting proposed general statements about regularities or phenomena in the social worlds and then seeking to verify them using evidence from empirical data. With analytic induction one begins by formulating a tentative hypothesis on the basis of an inspection of the data and then studies a small section of data to see whether the latter relates to it; if not the tentative hypothesis is reformulated or the phenomenon is redefined to exclude the case. The progressive modification of the hypothesis is guided by developing theoretical ideas and continues until there are no more negative cases to be found and a universal relationship is thus demonstrated between the hypothesis and the data. Znaniecki stated that the most common method used by natural scientists, that of the experiment, is not appropriate for examining the nature of social facts but claims the **logic** of the experiment can be incorporated into social research and the necessary closed systems do not have to be created artificially because they occur naturally in the social world (Hammersley, 1989:166).

There are very few true examples of analytic induction, the most notable being those of Lindesmith's study of opiate addiction (1947), Cressey's work on financial trust violation (1950) and Becker's study of marijuana use (Becker, 1953). Lindesmith stated that he used the method of science and the logic of the experiment to discover universal causal laws (i.e. the necessary and sufficient conditions under which opiate addiction occurs). His initial hypothesis was that addiction occurs when a person knows what drug they are taking and has been taking it for sufficient time to cause withdrawal distress on removal. This hypothesis soon had to be reformulated because of the emergence of negative cases;[107] redefinition of phenomena and reformulation of the hypothesis continued until there were no remaining negative cases and he reached the position of being able to able to establish his theory that addiction occurs when one is "using the drug for the consciously understood purpose of alleviating the withdrawal symptoms" (Lindesmith, 1937:3 cited in Hammersley, 1989:168). Lindesmith did, however, recognise that the emergence of new negative cases in the future would necessitate a revision of the hypothesis.

Cressey's work provides another example of the practice of analytic induction. He began with an initial hypothesis which stated that "positions of financial trust are violated when the incumbent has learned, in connection with the business or profession in which he is employed, that some forms of trust violation are merely 'technical violations' and are not really 'illegal' or 'wrong'" (Cressey, 1950:741 cited in Hammersley, 1989:169). After the identification of negative cases and subsequent reformulations, he arrived at his final theory, stating that "trusted persons become trust violators when they conceive of themselves as having a financial problem which is non-shareable, have the knowledge or awareness that this problem can be secretly resolved by violation of the position of financial trust and are able to apply to their own conduct in that situation verbalisations that enable them to adjust their conceptions of themselves as trusted persons with their conceptions of themselves as users of the entrusted funds or property" (Cressey, 1950:742 cited in Hammersley, 1989:172).

Both Robinson (1951) and Turner (1969) have mounted challenges to Znaniecki's insistence on the essential separation of analytic and enumerative induction. Robinson (1951) demonstrated that, contrary to its claims, analytic induction provides only the **necessary** and not the **sufficient** conditions

[107] See section 5.3.4.

for a phenomenon to be explained because it demands that only the cases where the phenomenon occurs are studied, not the cases where it does not occur. In order to achieve the latter, one has to employ methods of enumerative induction, which Znaniecki had deemed to be inappropriate for researching the social world. Both Lindesmith and Cressey in fact went on to use the comparative (i.e. enumerative) method in order to determine whether there were any cases where the set of necessary conditions as specified by the hypothesis occurred but the phenomenon (i.e. sufficient conditions) did not exist. Robinson contested Znaniecki's rejection of enumerative induction on the grounds that its logic of searching for negative cases is identical to that of analytic induction:

> The qualitative contrast which Znaniecki sets up between analytic and enumerative induction as methods of causal analysis is thus only a quantitative contrast and is not basic. The difference is in how far you push your study before you publish your results. (Robinson, 1951:203.)

According to Znaniecki, analytic induction aims to provide perfect explanations for social phenomena, which in turn lead to the statement of universal laws. Robinson, however, claimed it is not necessary to discover a perfect correlation between cause and effect but that one should develop theories in probabilistic terms, while Hammersley asserted that evidence of less than perfect correlations can be used to support the discovery of universal laws for the social world because exceptions are bound to arise "from extraneous variables which we have not been able to control" (Hammersley, 1989:196).

Turner (1969) also demonstrated the limitations of analytic induction by exposing the impossibility of making empirical predictions from the universal statements which it generates. He uses the Lindesmith and Cressey studies to demonstrate that empirical prediction is unattainable for two main reasons, the first being that "there is no basis for determining beforehand whether the conditions specified as necessary will exist in a particular instance" (208). For example, in Lindesmith's study there is no basis for knowing who will use the drug, what the range of withdrawal symptoms will be or whether they might be so mild that they do not lead to addiction, who may acknowledge the symptoms and recognise how to get relief, and what potential effect personal and social factors might have on outcome. Turner claimed empirical prediction is also impossible because "the alleged preconditions or essential causes of the phenomenon under examination cannot be fully specified apart from observation of the condition they are supposed to produce" (208). He questioned whether, in Cressey's work, it is possible to claim that a problem is non-shareable until a person becomes involved in embezzlement in order to try to solve it. "If a man has not revealed his problem to others today, can we say he will not share it tomorrow?" (208–9).

Turner conceded that analytic induction does provide definitions, but believes it is difficult for the researcher to move beyond this stage because of the method's dependency on a closed system. This can only be activated from outside by what Turner terms an **intrusive factor** and moreover:

> **External variables operating upon any closed system do not have a uniform effect because they have to be assimilated to the receiving system in order to become effective as causes.** The outside variable has to be translated, in a sense, into a cause relevant to the receiving system. Normally there will be alternate ways in which the same external variable may be translated depending upon the full context within which it is operative. The situation in which a man finds himself, for example, can only activate the closed system of the embezzlement process when it becomes translated into a non-sharable problem. (213 original emphasis)

Turner stated that there is always some doubt about the ways in which intrusive factors may, or indeed if they will, activate a system and therefore empirical prediction will always demand some form of probability statement. He suggested that if the methods of analytic and enumerative induction are regarded as being complementary rather than antithetical, then one can move towards a position of being able to use the closed system for empirical prediction because it can provide indications of correlations that are worthy of testing, and from this "probability associations can be organized into meaningful patterns" (216).

Although the original claim of analytic induction was that it would lead to perfect universal explanations, Katz (1983) stressed we should not be discouraged by the method's failure to realise this goal. He contested the idea that analytic induction is superior to enumerative induction as a method for investigating the social world but, having used analytic induction to assess the careers of legal assistance lawyers, he asserted it should be regarded simply as a **procedure**:

> Analytic induction's quest for the perfect explanation or 'universals' should be understood as a strategy for research rather than as the ultimate measure of the method. Analytic induction is a method for conducting social research, not a perspective from which to evaluate findings. (133)

Katz was not attempting to devalue the method, but stressed its logic can still be effective for other purposes: "The test is not whether a final state of perfect explanation has been achieved but the **distance** that has been travelled over negative cases and through consequent qualifications from an initial state of knowledge" (133).

There are several well known examples of qualitative health studies which have used the logic of analytic induction, but the researchers involved are often keen to stress the areas in which their aims and practice differ from those of Znaniecki's classic approach. In his study of surgeons' decision making in adeno-tonsillectomy cases, Bloor (1976)[108] attempted to reconstruct each specialist's decision rules for deciding whether to perform operations in order to examine significant variations in assessment. Observations were carried out at the out-patient consultations of 11 specialists. Bloor described his method of analysis as "a modified analytic inductive technique", which was comprised of the following seven stages.

1. For each specialist separately, cases were provisionally classified according to the disposal category into which they fell.

2. The data on all a specialist's cases in a particular disposal category were scrutinised in order to attempt to produce a provisional list of those case-features common to the cases in that category.

3. The **deviant cases** (i.e. those cases where features common to many of the cases in the disposal category were lacking) were scrutinised in order to ascertain whether (a) the provisional list of case features common to a particular category could be so modified as to allow the inclusion of the deviant cases; or (b) the classificatory system could be so modified as to allow the inclusion of the deviant cases within a modified category.

4. Having thus produced a list of case features common to all cases in a particular category, cases in alternative categories were scrutinised to discover which case-features were shared with cases outside the first category considered. Such shared case-features were thus judged necessary rather than sufficient for the achievement of a particular disposal.

5. From the necessary and sufficient case-features associated with particular category of cases sharing a common disposal, the specialist's relevant decision rules were derived.

6. For each decision rule the cases to which it applied were rescrutinised to derive the search procedures associated with that decision rule.

7. The above steps were then re-enacted for each of the other specialist's disposal-categories until a set of decision rules and associated search procedures for that specialist was derived, which accounted for the disposal of all the specialist's cases for which data had been gathered. The analysis was then repeated for each of the remaining specialists in turn (Bloor, 1976:45–46).

From this analysis, Bloor identified five parameters of variation in the specialists' routines, which he saw as being the cause of systematic variation in patient assessment. These concerned the search procedures used in examination of the child, the decision rules relating to the examination findings, the search procedures used in the history-taking, the decision rules pertaining to the history and the routines used according to the age of the child. Bloor stated that character of medical knowledge itself could be responsible for these differences because "disease entity is essentially arbitrary" (55) but also drew on the work of the eighteenth century philosopher Bishop George Berkeley[109] in order to explain his findings. Berkeley argued that men's conceptions are not framed as abstract ideas, but are detailed and explicit. Therefore, Bloor claimed, "the various routines of each specialist can be seen as referring to a series of **specific** ideas of those signs, symptoms and circumstances for which surgery is indicated, with each specialist differing from his colleagues as to the nature of those specific ideas" (45 original emphasis).

He concluded that there are three main types of variation in the specialists' judgements of appropriate outcomes in this context. The first two, namely "variation between specialists in the specificity of their decision rules pertaining to a given class of case" and "difference in the specificity of the specialist's decision rules (a difference in the amount of error that will be tolerated, a difference in the specialist's conception of his purpose at hand) between different classes of cases", may primarily offer an explanation for differences in specialist

[108] See appendix 1 for details of this study; [109] See section 2.2.3 for a discussion of Berkeley's philosophy.

assessment, whereas the third "variation in the specificity of ideas between different settings or contexts" may be responsible for their persistence.

Bloor did not consider his approach to be "entirely inductive" (59).[110] He had, however, gained clear impressions during the data collection phase which had been fed into the analysis process. He also emphasised that this work does not aim to discover any form of universal explanation for the specialists' behaviour but that the framework offered him a way of "**systematically** organising an enormous quantity of data" (59 original emphasis). Bryman (1988) pointed out that Bloor did not formulate an explicit hypothesis immediately after providing a rough definition of the problem under study, neither did he interrupt his research in order to revise his theoretical ideas once a deviant case appeared because the analysis was not carried out until after the data collection was completed, but (as has been outlined above) this is unusual in qualitative research (83).

Sometimes researchers may combine the logic of analytic induction with other methods of analysis. In a study of paediatric clinics and wards, Strong and his co-researcher carried out observations of 1200 consultations between staff, parents and children or staff and children over a three and a half year period (Strong, 1979a).[111] Over a thousand of the consultations took place in a Scottish city but later fieldwork was carried out in a comparable North American location. Strong's aim was to determine what type of social occasion (6) each consultation represented and he identified the bureaucratic format, a standard type of consultation, but compared this with private, clinical and charity formats.

There was no initial intention to examine the doctor–patient role relationship in this work, yet it became clear this was significant as the research progressed. Although the study design was already well established, by using a combination of analytic induction and the **constant comparative method** (see below), Strong was able to generate a set of propositions about the doctor–patient role relationship and then test them. He began by using the constant comparative method on half of his data to see if it fitted the existing categories. Once he had developed "a more sophisticated version of a category" (235) he only coded if the material indicated a new feature of the category or where the category was of great theoretical significance but there were very few examples. In cases where the data did not fit any of the categories, then they were modified or new ones were constructed. Strong was then able to generate a set of propositions which he tested by analysing the second half of the data using the method of analytic induction, revising his hypotheses where the data did not fit the argument.

As explained above, the method of analytic induction is based on the search for, and review of, negative or deviant cases which enables the researcher to refine theoretical statements by establishing their limits.[112] In a study of consultations in a paediatric cardiology clinic, Silverman (1981) identified Down's syndrome children as negative cases because the recommended course of action for such patients was to adopt a policy of non-intervention, even where the condition was immediately life-threatening. "Whereas a 'normal' child with suspected congenital heart disease would be put up for catheterisation and, where clinically appropriate, for surgery, a Down's child with an identical cardiac lesion would usually receive neither catheterisation nor surgery" (254).

Like Bloor, although he used the logic of analytic induction, Silverman was keen to stress that he did not claim his findings to be universal; different clinics operate different policies, hence "geography can have an undesirable impact on the extent and nature of treatment" (272).

As discussed elsewhere in this report, qualitative research can be used to develop existing theory and project findings can be applied in subsequent investigations.[113] It is also useful to apply the logic of analytic induction across from one piece of research to another in order to build cumulative knowledge. For example, Dingwall and Murray's study of the treatment of children in A&E departments (Dingwall and Murray, 1983[114]) built on Jeffrey's categorisation of patients (Jeffrey, 1979). Jeffrey had concluded from his 7-month participant observation study of three adult accident departments in an English city that medical staff broadly classify their patients as belonging to one of two main categories. 'Good' patients fitted three criteria.

- They allowed medical staff to practise skills necessary for passing professional examinations.

[110] See section 3.2.1.2 for a discussion of the logics of induction and deduction; [111] See appendix 1 for details of this study; [112] This is discussed more fully in section 5.3.4; [113] See section 3.2.1.2; [114] See appendix 1 for details of this study.

- They allowed staff to practise their chosen speciality.
- They tested the general competence and maturity of the staff.

'Bad' patients, or 'rubbish' in staff argot, were mostly of four kinds: trivia, drunks, overdoses, and tramps. Jeffery argued that they break one or more of four rules.

- Patients must not be responsible for either their own illness or for getting better: medical staff can only be held responsible if they are able to treat the illness.
- Patients should be restricted in their reasonable activities by the illnesses they report.
- Patients should see illness as an undesirable state.
- Patients should cooperate with the competent agencies in trying to get well. (Dingwall and Murray, 1983:130–1.)

Dingwall and Murray's work formed part of a larger study of decision making by agencies responsible for the care and protection of children who were thought to have been neglected or abused. Participant observation was carried out in four English accident departments and interviews were conducted with medical and nursing staff; staff from elsewhere in the health service and from personal social services were also interviewed. Their data showed that:

- The vast majority of injuries sustained by children directly result from their own behaviour.

  > We observed cases of children falling off bicycles, ponies, donkeys, swings and slides, out of trees and downstairs. Children had pulled hot tea over themselves, ingested tablets, and cut themselves on a variety of objects. (133)

- Most of the injuries did not restrict the child's normal activities.
- The children's responses were not usually consistent with their injuries; some clamoured for attention over minor injuries, yet some of those who had sustained much more serious damage made very little fuss.
- The children were often uncooperative.

  > They frequently refuse to be examined, protest when needles are stuck in them and reject attempts to insert stitches or apply plaster. (133)

It became clear, therefore, that using Jeffrey's categories as an initial hypothesis, children were a negative case. They broke Jeffery's four rules which, in the case of an adult, would lead to categorisation as 'bad' or 'rubbish' but the medical staff did not punish the children in the same way. Dingwall and Murray then went on to develop a more sophisticated model of professional decision making on the basis of their findings.

4.6.1.4 Grounded theory

> A grounded theory is one that is inductively derived from the study of the phenomena it represents. (Strauss and Corbin, 1990:23.)

The grounded theory approach was originally devised by Glaser and Strauss in the early 1960s during research into hospital staffs' management of dying patients. It represents an elaboration and operationalisation of analytic induction;[115] with this method the researcher begins with an area of study, and the knowledge which is relevant to the area is allowed to emerge by way of the analysis process. Grounded theory is similar to analytic induction in that data are collected and after a general reflection on 'issues of concern' categories which fit the data are generated (Bryman and Burgess, 1994:4).

Grounded theory was formulated by Glaser and Strauss in reaction to contemporary ways of trying to establish closer links between sociological theory and research.

> Previous books on methods of social research have focused mainly on how to verify theories. This suggests an overemphasis in current sociology on the verification of theory and a resultant de-emphasis on the prior step of discovering what concepts and hypotheses are relevant for the area that one wishes to research. Testing theory is, of course, also a basic task confronting sociology. We would all agree that in social research generating theory goes hand in hand with verifying it; but many sociologists have been diverted from this truism in their zeal to test either existing theories or a theory that they have barely started to generate. (Glaser and Strauss, 1967:2.)

Glaser and Strauss challenged the dominance of quantitative analysis methods. They disagreed with the practice of many quantitative researchers who, being more concerned with hypothesis testing, tried to force their data into predefined categories and were afraid follow up unexpected emerging ideas for fear of undermining the rigour of their work. They also wanted to introduce sociologists to a strategy that would discourage them from the misuse of theory.

[115] See section 4.6.1.3.

> Grounded theory can help to forestall the opportunistic use of theories that have dubious fit and working capacity. So often in journals we read a highly empirical study which at its conclusion has a tacked-on explanation taken from a logically deduced theory. The author tries to give his data a more general sociological meaning, as well as to account for or interpret what he found. He uses this strategy because he has not been trained to generate a theory from the data he is reporting so that it will help interpret or explain the data in a general manner. (Glaser and Strauss, 1967:4.)

The methods outlined by Glaser and Strauss can be used to develop two basic types of theory. **Substantive theory** is generated in order to explain elements of an empirical area of sociological investigation "such as patient care, race relations, professional education, delinquency or research organisations", whereas **formal theory** relates to a conceptual area "such as stigma, deviant behaviour, formal organisation, socialisation, status congruency, authority and power, reward systems, or social mobility" (32). Substantive and formal theories are both "middle-range" theories which "fall between the 'minor working hypotheses' of everyday life and the 'all-inclusive grand theories'" (33). Glaser and Strauss stressed both of these types of theory must be grounded in data. They criticised the practice of merely "applying a few ideas from an established formal theory to a substantive area" as this would never generate valid substantive theory for any empirical situation (33).

Glaser and Strauss regarded two main techniques as being central to grounded theorising.

- The **constant comparative method** is a means of suggesting, but not testing, properties and hypotheses about general phenomena. Unlike analytic induction,[116] the constant comparative method does not claim to discover universal truths, or to offer proof of causes. During analysis **memos** are written, which serve as a record of the whole analytic process. Memos contain products of the actual coding, as well as descriptions of the theoretical sensitising and decision making which has occurred at each part of the process.

The constant comparative method is comprised of four stages.

1. Incidents in the data are coded in as many categories of analysis as possible, whilst at the same time comparing them with previous incidents coded in the same category. This will lead to the generation of theoretical properties of the category. For example, in Glaser and Strauss's own work (Glaser and Strauss, 1965b), they had constructed the category of 'social loss' of dying patients, but it soon became clear that this category encompassed a range of observations from a high to low degree of social loss, depending on the status of the patient (106).

2. The researcher then moves on to develop theories by integrating the categories and their properties. For example, by comparing incidents in the same study it was found that nurses constantly recalculated the degree of social loss of patients as they found out more about them, and that this in turn affected the way in which they maintained their professional composure whilst attending to the patients (109).

3. De-limiting of both the theory and the categories then takes place. Uniformities in the original set of categories or their properties are identified and then the theory can be formulated using a smaller set of higher level concepts. Glaser and Strauss moved from a point where they were considering all of the individual strategies used by nurses to cope with their dying patients to looking at **loss rationales** (110). After this has been achieved, the list of categories can be cut down. After a number of incidents have been coded into the same category, the researcher will only code the next incident if it points to some new aspect to the understanding of the category. If nothing new is being revealed, "it only adds bulk to the coded data and nothing to the theory" (111). At this point **theoretical saturation** is said to have been achieved. Theoretical saturation also overcomes the problem of deciding whether to recode all the data when a new category emerges half way through the coding process; if one continues and the new category also becomes theoretically saturated, this indicates nothing has been lost by not having used the code on the first half of the data.

4. When the coding process is completed, the researcher will have a set of coded data, memos (see above) and a theory. Providing he is "convinced that his analytic framework forms a systematic, substantive theory, that it is a reasonably accurate statement of the matters studied, and that it is couched in a form that others going

[116] See section 4.6.1.3.

into the same field could use" (113) he will be ready to produce a report of his findings.

- Categories may be elaborated by collecting further data through a process known as **theoretical sampling**[117] which involves sampling on the basis of concepts that have proved to be theoretically relevant to the emerging theory. The researcher selects further groups or subgroups from which to collect more data in order to generate as many properties of his categories as possible. If (s)he is engaged in generating substantive theory, (s)he selects groups of a similar substantive class in any setting, for example, a comparison of an emergency ward to all kinds of medical wards in any hospital in any country (53). In order to develop formal theory, (s)he will choose dissimilar substantive groups for comparison, such as an emergency ward and a fire department as examples of emergency services (54). Glaser and Strauss set out a clear explanation of the importance of theoretical sampling:

  > Why does the researcher's comparison of groups make the content of the data more theoretically relevant than when he merely selects and compares data? The answer is threefold. Comparison groups provide, as just noted, control over the two scales of generality: first conceptual level, and second, population scope. Third, comparison groups also provide simultaneous maximization or minimization of both the differences and the similarities of data that bear on the categories of data being studied. This control over similarities and differences is vital for discovering categories and for developing and relating their theoretical properties, all necessary for the further development of an emergent theory (55).

Theoretical sampling continues until the categories become saturated (see above). It is at this point that the theory can be considered to be conceptually adequate.

The publication of *The Discovery of Grounded Theory* (Glaser and Strauss, 1967) represented a historical milestone in thinking about qualitative data analysis. Glaser and Strauss are still frequently cited as being the authorities on this method but it is probably now more appropriate to think about grounded theory as being a set of approaches for the generating and advancing of theory, rather than as a specific, single technique.

The grounded theory method has been subsequently refined by Strauss (1987) and Strauss and Corbin (1990). Whilst Strauss and Corbin presented their method as grounded theory, this has been strongly challenged by Glaser (1992). He asserted that Strauss and Corbin's book *Basics of Qualitative Research* "distorts and misconceives grounded theory, while engaging in a gross neglect of 90% of its important ideas" (2). A recent paper by Melia (1996) detailed Glaser's objections to Strauss's new version of the method. She described Glaser's claims that Strauss and Corbin have outlined the technique of **full conceptual description**, which he regards as a fundamentally different type of analytical approach because "there is a forcing of data and a preconceived and verificational approach to qualitative data analysis" (3). In other words, the approach is based on strategies that give rise to the very problems Glaser and Strauss originally set out to eradicate. Glaser challenged many other elements of Strauss and Corbin's work. For example, they claimed a research question in a grounded theory study is a statement which identifies the phenomenon to be studied, yet Glaser restated that, according to their original work, phenomena in grounded theory emerge from coding, theoretical sampling and constant comparison during the analysis phase. There is also a discrepancy between Strauss and Corbin's notion of initial conceptualisation of data in which each item is broken down and coded separately, and Glaser's definition of conceptualisation in which incidents and/or concepts are compared and categories are built up until they become saturated. Strauss and Corbin made no reference to saturation, yet this was one of the main elements of Glaser and Strauss's earlier work (Melia, 1996).

In spite of this dispute, grounded theory remains the most popular formal method of qualitative data analysis. However, researchers have tended to develop their own styles of the grounded theory approach and it is necessary to ensure that the term grounded theory is not merely being used as that which Richards and Richards (1991) called "an approving bumper sticker" in reports of qualitative studies (43). Having outlined the main components of grounded theory, Bryman (1988) issued this warning:

> In spite of the frequency with which Glaser and Strauss and the idea of grounded theory are cited in the literature, there are comparatively few instances of its application along the lines developed above. The term is often used as a way of conveying the notion of an approach to the generation of theory which is

[117] See section 4.1.1.4 for a discussion of theoretical sampling.

> derived from a predominately qualitative research base. Much qualitative research relies on the elucidation of a theoretical framework subsequent (rather than during) the data collection phase. The idea of grounded theory is often used as a way of justifying the use of a qualitative research approach i.e. so the work can be confirmed as respectable. (85)

Researchers who use the grounded theory method can be expected, at a minimum, to incorporate the following main strategies.

- The structure of the inquiry will be fundamentally shaped by the aim to discover social and social psychological processes.
- The data collection and analysis phases of the project will proceed simultaneously.
- The analytic processes employed will prompt theory discovery and development rather than the verification of pre-existing theories.
- Theoretical sampling will refine, elaborate and exhaust conceptual categories.
- Systematic application of grounded theory analytic methods will progressively lead to more abstract analytic levels. (Charmaz, 1983b:125.)

Bryman (1988) cited Hammersley's suggestion that the time consuming practicalities of qualitative fieldwork such as tape recording and transcription "may render the grounded theory framework of a constant interweaving of categories and data almost impossible to accomplish" (85). There is a dearth of examples of true grounded theory studies; however, there are cases where research has been carried out using the principles of grounded theory analysis strategy.

Charmaz's study of loss of self amongst the chronically ill provides a clear example. She stressed that grounded theorists are primarily concerned with generating and refining theory. They do not rely on existing literature to shape their ideas or "pour their data into someone else's theoretical framework or substantive analysis" because this stifles innovation and may serve to "perpetuate ideas that could be further refined transcended or discarded" (1983b:110–11). Data were collected in Northern California in the form of 73 in-depth interviews with 57 chronically ill people who were suffering from conditions such as cardiovascular disease, diabetes, cancer and multiple sclerosis (1983a). During the initial coding phase (see above), in order to summarise the large amounts of information and also to establish what was emerging from the data, Charmaz undertook four activities:

- the collection of information about "the general context, central participants and their roles, timing and structuring of events and the relative emphasis participants place on various issues in the data" and tried to establish "connections between individuals' special situations and problems and their interpretations of their experience"
- the development of codes that drew attention to missing aspects of information, for example, patients' lack of knowledge about their condition
- the searching for *in vivo* codes in the data where codes are constructed using elements of the interviewees' own speech and description. This led to the adoption and development of concepts such as 'super-normal' and 'being there' within the coding framework
- the coding of feelings which may be implicit in the talk of some respondents (1983b:114–15).

A wide range of diverse codes, such as self-perception, awareness of difference, identifying self through ill-health and self esteem, were thus generated as a result of the initial coding process and this allowed Charmaz to pursue many lines of inquiry from her data. Subsequently, during the focused coding process, she applied the initial set of codes to large amounts of data and raised the classification of the data to a more analytical level. Categories were developed and woven into a **processual analysis** through which she could "abstract and explicate experience" (1983b:117). For example, she stated: "... after categorising types of self pity and its social sources, I then developed the processual categories of becoming immersed in self-pity and reversing self-pity which were vivid when I directed questions towards them, but only implicitly related in the early data before I systematically explored these areas" (118).

As a result of the use of the grounded theory analysis strategy, Charmaz was able to suggest that, according to her respondents, sources of suffering loss of self among the chronically ill were the necessity of having to lead restricted lives, the experience of social isolation, being discredited and becoming a burden to others (1983a:170).[118]

[118] Further examples of the use of grounded theory in health-related research can be found in a recent volume edited by Strauss and Corbin (1997).

4.6.1.5 Cumulative research

Hammersley, Scarth and Webb (1985) stressed the importance of cumulative research, an analytical method which combines the features of both hypothesis advancement and testing. It also offers a way of moving on from the potentially difficult problem faced by those involved in qualitative data analysis, namely that if theory is formulated prior to empirical investigation, it can restrict the analysis process, yet if one leaves theory elaboration until fieldwork is underway, one may encounter data management problems, the difficulties of trying to hold theoretical considerations in abeyance and the problem of how to choose a suitable research site in the first place (Bryman, 1988:87).

Hammersley, Scarth and Webb stated that although descriptions and explanations from individual study findings can be pieced together in order to build up a composite picture (see above), "the logic of developing and testing theory is quite different" (54) because here one is concerned with establishing which theory is correct. Because it is usually impossible to achieve this by carrying out a single study, multi-site research and the drawing of comparisons between different projects becomes important. They cited the ethnographic school studies previously undertaken by Hargreaves, Lacey and Ball which were conducted at a different research sites but all examined differentiation-polarisation theory. This claimed that "where pupils are differentiated according to a set of academic and behavioural values, pupils will become polarized in their attitude to those values, with those pupils given a low rating rejecting the values and behaving accordingly" (54). Hammersley, Scarth and Webb regarded this comparative method of analysis as "the key to the development and testing of theory" (56) and emphasised its connection with both analytic induction and grounded theory where cases are also analysed in relation to each other from a theoretical perspective.

In their application of the comparative method, Hammersley, Scarth and Webb went on to adopt the elements of theory development from grounded theory, the interrelation of theory development and testing from analytic induction and the importance of critical case selection for facilitating theory testing from the work of Hargreaves, Lacey and Ball. Here Hammersley, Scarth and Webb set out to examine their own hypothesis that "external examinations lead to lecturing and note-taking on the part of secondary school teachers and to rote-learning and instrumental attitudes among their pupils" (58). They concluded that the comparative method is "the best hope for the cumulative development of sociological knowledge" (64).

4.6.2 The use of software in qualitative data analysis

It has become standard practice for contemporary texts on qualitative research to include chapters on the use of computers (Dey, 1993; Richards and Richards, 1994b; Durkin, 1997). This reflects the relatively recent development of software packages designed to assist the data analysis process. During the 1960s social scientists made increasing use of statistical programs on mainframe computers when working with quantitative materials, but qualitative researchers had to rely on traditional manual techniques which involved the cutting and pasting together of material from printed hard copies of their data. It was not until the advent of microcomputers in the early 1980s that qualitative data analysis packages were initially introduced with the adaptation of basic programs for textual analysis for use on MS-DOS machines. Towards the end of the 1980s programs for Macintosh computers were also developed and subsequently a wide range of qualitative data analysis software has become available.

Packages that have been designed specifically for the analysis of qualitative data have largely resulted from the efforts of practising academics, for example, John Seidel (developer of The Ethnograph) and Tom and Lyn Richards (developers of NUD*IST) who have become involved in program development in order to serve both their own needs and those of the research community. Thus, those involved in the technical practicalities of software development have also been in a position to engage in the debates which have emerged amongst researchers over the implications of the growing use of computer packages for the theory and practice of the analysis process.

4.6.2.1 The advantages and disadvantages of the use of CAQDAS

Practical issues. Many people hoped the introduction of computer software for the analysis of qualitative research would lead to the development of more effective working practices and efficient data management. The prospect of handling larger data sets and being able to organise material within a shorter space of time was attractive, particularly as this would allow more energy to be spent on intellectual and creative thinking. (Conrad and Reinharz, 1984; Tesch, 1990.) Prior to this point,

qualitative analysis had usually been conducted by independent, fallible researchers; now both team work and multi-site projects would be facilitated and the computer could be employed in the role of a clerk "with three exceptional capabilities: a perfect memory, the ability to retrieve any document immediately and the capacities of an untiring and very fast typist" (Gerson, 1984:64).

However, early commentators warned of potential problems concerning loss of information through operator error or computer failure and breach of confidentiality as more individuals could gain access to data sets (Conrad and Reinharz, 1984). More recently, as the range of qualitative data analysis packages has expanded, academics have also been alerted to potential problems concerning the selection, purchase and implementation of analysis programs. It is critical that a package is chosen carefully in accordance with the practical and theoretical needs of a specific project,[119] yet availability of software and financial restrictions may well prove to be influential.

Lee and Fielding (1995) have outlined the way in which many researchers come to hear about programs through informal networks as well as conferences, workshops and seminars, but the recent publication of key texts which present summaries of currently available software has also helped to inform the decision-making process (see below). If researchers have no prior experience with the particular program which has been selected, it is essential they learn how to use it correctly before beginning to work with their data in order to avoid eventual problems of producing **off-target** analyses. As Lee and Fielding have warned, it is the misapplication of a package that can cause problems – not the package itself (Lee and Fielding, 1991).

There are other areas where caution must be exercised. Qualitative data analysis packages may offer the attractive inducement of a speedier completion of the research process, but Lee and Fielding also call for resisting the temptation to indulge in what they term as "quick and dirty research", with "its attendant danger of premature theoretical closure" (8). The opportunity to work with larger data sets can also present problems in that they take longer to collect and prepare (Kelle and Laurie, 1995). Seidel (1991:107–16) has feared researchers' obsession with volume will lead to what he terms as 'analytic madness' where the component of quantity will be elevated whilst that of quality is sacrificed.

Issues connected with the coding and analysis process. Many have questioned the wisdom of using computers in qualitative data analysis as "the central analytic task in qualitative research – the understanding of the meanings of texts – cannot be computerised because it is not an algorithmic process and hence cannot be considered a mechanical task" (Kelle, 1995:3). However ease of coding by using software has proved to be attractive. The majority of current CAQDAS programs offer facilities for coding and retrieval[120] and there is little doubt that the mechanics of assigning codes to qualitative data become much more straightforward if a software package is used. Thinking about the meaning of the data set and the simultaneous manipulation of its components is much more readily achieved than is the case with manual analysis. It is very easy to add extra codes or alter existing ones and, as a result, the growth of a coding or indexing system becomes a much more intrinsic part of the theoretical development of a study, although one must guard against creating a situation of **methodological anomie** because one cannot code for ever! (Richards and Richards, 1991).

Lonkila (1995) claimed that the development of grounded theory is facilitated by some types of qualitative data analysis software because comparisons can be made easily within data sets, thus producing a more methodical advancement of hypotheses and concepts. Connections between pieces of data can also be indicated easily through the use of axial coding[121] between codes and text segments and between codes themselves, thus assisting the theory building process.[122]

However, it appears the introduction of technology into the coding process has also had some negative effects. Seidel (1991:112) has referred to the epistemological problem of the possible reification of the relationship between researchers and their data. As coding becomes easier through the use of qualitative data software packages, he feared the researcher might come to believe the phenomena being coded are actually out there in the world because they can be labelled, whereas, he stated, they are more like artifacts, resulting from the relationship the researcher has with the data and the intellectual baggage (s)he brings to a study.

[119] See section 4.6.2.1; [120] See section 4.6.2.1; [121] See section 4.6; [122] See section 4.6.2.1.

Weaver and Atkinson (1994) have commented that the neat retrieval of data may lead the researcher into a false sense of security over the analysis because (s)he is removed from the original messiness of the piles of paper, scissors and glue which has always been the hallmark of qualitative work. Indeed, developers such as John Seidel have been explicit in their beliefs concerning the desirability for qualitative researchers to remain close to, and immersed in, their data, yet some workers have reported problems of feeling distanced. This is not only because of having to work on screen as is the case with several packages, but also because the use of computers enables easier working with larger data sets and therefore more people are often employed to be responsible for transcription and data entry, thus the researcher loses initial opportunities to familiarise him or herself with the content of the data (Lee and Fielding, 1995).

In addition, Weaver and Atkinson (1994) have described the apparently fundamental mismatch between the primarily holistic aims of much of qualitative research and the coding strategies employed by many of the CAQDAS packages, the problem being that as the data is increasingly split up into chunks and coded, the original context is lost. They argued that chunking and coding also tends to discourage reflexivity, another hallmark of qualitative research. However, it is hoped the recent development of hypertext[123] should work towards eliminating this difficulty because it facilitates a much more interactive relationship between the researcher and the source material (Coffey *et al,* 1996).

There are debates surrounding the aims of coding using data analysis software. Technology may have made the mechanical process easier, but qualitative researchers must remain clear about the purpose of their actions during the coding procedure, as the computer is not capable of making informed decisions. Weaver and Atkinson (1994) cited the claim of Miles and Huberman (1994) that coding of data is an intrinsic part of analysis because one's knowledge of a subject grows through the coding process. Likewise, Richards and Richards (1994b) have argued coding itself forms an integral part of theory emergence and have disputed the current tendency to sub-divide qualitative data analysis packages into those designed for code and retrieval and others, which claim to support theory building.

Richards and Richards (1991), Weaver and Atkinson (1994) and Lonkila (1995) have commented on emerging fears that the potential dominance of a few popular packages may well serve to standardise the outcome of qualitative studies instead of expediting individuality. Because certain packages such as NUD*IST, ATLAS/ti Kwalitan and HyperRESEARCH are based on a specific way of analysing, namely grounded theory,[124] then this method may well become dominant as a consequence of increased software use. Although Richards and Richards (1991) called for an active reversal of this potential trend, urging program developers to seek to "celebrate the diversity" of qualitative methods, others have suggested the use of computers in this type of data analysis could begin to drive the research process and pose a threat to the very essence of qualitative studies – "the computer had shifted in my worst case scenario from an aid in doing ethnography to a definition of what ethnography might do" (Agar, 1991). In spite of Pfaffenberger's conviction that technical determinism should not be a problem because both the software and the machines are constructed in a social setting to which users bring their own knowledge (Pfaffenberger, 1988), this is a concern that has been taken up in ensuing theoretical debates and is one of which all practising qualitative researchers should remain aware.

Issues of rigour, validity and reliability. The introduction of computer-assisted qualitative data analysis was accompanied by claims that the findings of qualitative research would now be recognised as being more comprehensive, trustworthy and rigorous (Conrad and Reinharz, 1984). 'Harder' analyses of 'soft' data would be promoted through the use of computers which could not only provide statistical-type summaries of texts such as in the form of word counts, but would also be able to offer a more exacting and robust method of employing qualitative techniques (Richards and Richards, 1991a).

Kelle and Laurie (1995) commented on the 'seductive' nature of the arguments to support increased rigour which had appealed to those who sought to contest the negative image of qualitative research. Its methods had previously been depicted as both unsystematic and subjective and had been compared unfavourably with those of quantitative investigation. However, others had dismissed this

[123] In hypertext programs, pieces of data are joined together in meaningful ways by electronic links thus enabling the reader to follow pathways or trails through the text; [124] See section 4.6.1.4 for a discussion of grounded theory.

possibility of change as being totally inappropriate for the qualitative model: "we don't do validity and reliability" (Smith 1984 cited in Kelle and Laurie, 1995:20)[125] and again it was feared the computer may be "a genie in the bottle which, once released, will transform the activity of field research in unnoticed and unwelcome ways" (Lee and Fielding, 1991:6).

Proposals to 'clean up' qualitative practices have been contested by those who refuse to accept quantitative methods as being the gold standard for all research. There have been concerns that if more emphasis was to be placed on these issues, then the use of computer software for data analysis would be serving to obscure the division between quantitative and qualitative research and that qualitative methods would be in danger of losing those particular attributes which had always made them distinctive. For example, Hesse-Biber (1995) argued there is little point in encouraging the introduction of something that is "akin to significance testing" into qualitative work where single occurrences of phenomena are often recognised as being equally important for some purposes as those which are found in multiple instances (32).

Attempts to increase validity and reliability in qualitative research also give rise to philosophical questions about the very nature of understanding. Some have suggested that if one accepts the existence of criteria against which knowledge can be measured, this implicitly involves rejecting the challenge to universal truth (Weaver and Atkinson, 1994). Kelle and Laurie (1995: 19-20) argued that the epistemological underpinnings of many strands of qualitative research will not allow the performance of validity and reliability exercises on its data.[126] For example, as naturalistic inquiry studies phenomena in natural settings, it is incongruous to attempt to validate such findings using models which come from experimental research.[127]

From a practical perspective, the facility of being able to work with larger data sets on a computer will not automatically lead to more valid and generalisable findings as the commissioning of larger sample does not guarantee a more genuine representation of the population. Kelle and Laurie (1995) also asserted that such strategies are fundamentally inappropriate for qualitative research because it tends involve theoretical rather than random sampling once phenomena have been discovered through in-depth analysis.[128]

4.6.2.2 Choosing a qualitative data analysis package

There are currently over 20 packages which have been designed to assist in the analysis of qualitative data. In order to be able to make a selection, it is important the researcher is clear about the type of project to be undertaken and what (s)he wants to achieve by using the package as "the logic of data analysis which is incorporated into a package may not necessarily be the same as that which is proposed by a certain methodology" (Kelle, 1995). The role of the computer has been misunderstood by some who thought the development of CAQDAS would offer a magical solution to the demanding task of qualitative analysis. As Weitzman and Miles (1994) reminded us: "There is no computer program that will 'analyse' your data ... Computers don't analyse data; people do" (3).

There are several publications that have outlined the currently available software packages. Tesch (1990) described the diversity of methods used by qualitative researchers and provided a clear summary of the software which had been developed to date to assist in the analysis of textual data. Ironically, Tesch has since been criticised for presenting her material in a way which reflects precisely what many fear will happen to qualitative work as a result of increased use of the computer. She has a tendency to focus on technical rather than epistemological issues, to be over-systematic in her synopsis which then fails to give a true representation of the messiness of qualitative work (Weaver and Atkinson, 1994). Nevertheless, this text gives a lucid description of the issues and concepts of qualitative research and, though some of the packages she described are now obsolete, it remains a useful resource.

Subsequent edited volumes by Fielding and Lee (1991) and Kelle (1995) both contain sections which provide brief summaries of currently available programs for qualitative data analysis. However, the most comprehensive and up-to-date publication in this field has been produced by Weitzman and Miles (1995). They have no specific

[125] See section 5.1 for a discussion of the criteria by which qualitative research might be assessed; [126] See section 5.2 for a discussion of this position; [127] These arguments are evaluated in section 3.2.1.3; [128] The merits of theoretical sampling and sampling for empirical generalisability are reviewed in section 4.1.

allegiance or commercial investment in any of the products and this book aims to provide the reader with clear, critical, unbiased information, which will assist in the selection of an appropriate package for use on a particular project.

Weitzman and Miles encouraged the reader to consider his or her own level of computer competence, the type of project which is to be undertaken and the kind of analysis which is anticipated. This work does not include packages reviewed by Tesch (1990), which were still available in the same unrevised form, or any new products which had become available after May 1994. The 24 programs under evaluation are grouped into five broad categories, which the authors describe as follows (5):

- **text retrievers**, which are designed to search for words or phrases in the dataset (Metamorph, Orbis, Sonar Professional, The Text Collector, WordCruncher, ZyINDEX)
- **textbase managers**, which sort and organise data (askSAM, Folio VIEWS, Tabletop, MAX)
- **code and retrieve programs**, which facilitate the application of codes or keywords to sections of data and then retrieve those sections by codes or combinations of codes (HyperQual2, Kwalitan, Martin, QUALPRO, The Ethnograph)
- **code-based theory builders**, which also have code and retrieve facilities but include the capacity to build conceptual structures and to formulate and test hypotheses (AQUAD, ATLAS/ti, HyperRESEARCH, NUD*IST QCA)
- **conceptual network builders**, which allow the researcher to use a network of nodes and links to assist in the formulation and representation of conceptual schemes (Inspiration, MECA, MetaDesign, SemNet).

Each program is reviewed separately, providing contact details of the developers, cost, hardware requirements, price, key features, requirements for data entry, an outline of user instructions, an assessment of user friendliness and a comparison with other packages that do similar things. This text provides critical information for all researchers, whether they are unfamiliar with the features of CAQDAS or they are experienced program users who need to ensure they choose the most apposite form of software before beginning a new project. Texts that describe formal assessment of programs, (e.g. Weaver and Atkinson, 1994; Stanley and Temple, 1995), and articles by researchers who have used computer packages in qualitative research projects (e.g. Mangabeira, 1995; Armstrong, 1995) are also available.

4.7 Ethics and qualitative research

Much discussion of the ethics of qualitative research focuses upon the appropriateness of applying ethical codes that originated in biomedicine to social scientific research in general, and qualitative research in particular. A number of authors (Cassell, 1978; Barnes, 1979; Cassell, 1979; Dingwall, 1980; Thorne, 1980; Walker, 1980; Wax, 1980; Cassell, 1982; Kelman, 1982; Finch, 1986; Merriam, 1988; House, 1990) have argued that the ethical issues raised by qualitative research cannot simply be subsumed within such biomedical codes.

Smith (1975) summarised the set of guidelines for ethical practice upon which major research and governmental institutions have come to agree, in order to regulate research involving human subjects.

- **Informed consent**
  The researcher is obligated to ensure that the subject understands what his participation in the study will involve, and the subject's consent to participate is obtained without coercion (13).
- **Confidentiality**
  The investigator should keep in confidence all information obtained about research subjects (14).
- **Subject rights and welfare**
  There should be awareness that investigators may, quite unintentionally, introduce unnecessary or unacceptable hazards, or fail to provide adequate safeguards (14).
- **Subject risk–potential benefits ratio**
  Known or foreseeable risks to subjects must be outweighed by the probable benefits that may accrue to them and/or humanity by their project participation (14).

These guidelines for ethical practice represent an attempt to operationalise some ethical principles which are held to be relevant to the practice of research that involves people. Beauchamp and co-workers (1982) identified four ethical principles around which ethical debates about acceptable research practices centre.

- **The principle of autonomy or self-determination**

  > The central demand constant in the diverse formulations [of autonomy] is that we respect the values and decisions of other people. (18)

  In relation to research which involves human beings, this principle is closely involved in the issue of informed consent (see above).

- **The principle of nonmaleficence**

  This principle stipulates that it is wrong intentionally to inflict harm on another person. (18)

- **The principle of beneficence**

  [Beneficence] specifies a positive obligation to remove existing harms and to confer benefits on others. (19)

  Together the two principles of nonmaleficence and beneficence inform the practice of risk–benefit analysis (see above).

- **The principle of justice**

  This principle stipulates that people who are equal in relevant respects should be treated equally. (19)

Beauchamp and co-workers (1982) also argued that ethical theories can be seen as falling into one of two distinct patterns (see also Brewster Smith, 1979). On the one hand, **utilitarian or consequentialist** ethics assess actions in terms of their outcomes or consequences. From this perspective, research practice is justified if its benefits outweigh any potential harm. This pattern lends itself particularly well to the combination of the principles of beneficence and non-maleficence in risk–benefit analysis (see above). They contrast this with **deontological or Kantian** ethical theories, which treat features of a research practice, other than its consequences, as relevant to a judgement about its morality. From a Kantian perspective, for example, a research practice is deemed unethical insofar as it treats people as means toward some extraneous ends rather than as ends in themselves (Kelman, 1982). Macklin (1982) offered the following summary of the Kantian position: "Right actions (or practices) are those that demonstrate respect for persons; in particular, they never treat people as means solely, they do not violate autonomy, and they prohibit exploitation" (194).

This perspective has important links with both the principle of autonomy and with the practice of obtaining informed consent.

We have, then, at least three levels of analysis in relation to ethical aspects of research practice:

- At the most abstract level, we have the distinction between deontological and consequentialist ethics. The debate about the appropriateness of applying ethical codes derived from biomedical research to social scientific and/or qualitative research is rarely conducted at this level. Both qualitative and quantitative researchers, as well as biomedical scientists, are to be found on both sides of that divide.
- At the next level of analysis, that of the ethical principles which should properly be applied to research practice, there is a general consensus that all four principles are 'a good thing'. Researchers from very different methodological traditions share a common commitment to autonomy, nonmaleficence, beneficence and justice. There is, at times, some disagreement about which principles should be prioritised in different kinds of research.
- It is, however, at the most concrete level of analysis, when an attempt is made to operationalise these principles in terms of research practice, that controversy most commonly arises.

Thus the debate about the appropriateness of imposing ethical practices, derived from biomedical research upon social science/qualitative research, centres upon whether or not biomedical and social science research are sufficiently similar to permit ethical practices originating in biomedical research simply to be lifted and mechanically applied to social science research.

4.7.1 Alternative ethical codes for social science and/or qualitative research?

The argument in favour of developing an alternative ethical code, which is specific to social scientific and/or qualitative research takes two, often complementary, forms.

- The mechanical application of ethical codes, developed in the context of biomedicine, may be unduly constraining in social scientific research, and may lead to the ritualistic and distorting observation of rules, at the expense of genuine ethical practice.
- A focus on ethical concerns, derived from the practice of biomedical research, may in fact serve to distract attention from those risks that are specific to qualitative research. Cassell (1979) put this case forcefully:

  Can regulations which conceptualize the research enterprise in a way which is inappropriate for a particular type of research protect those who are studied by this technique? ... I can imagine a definite temptation for an investigator to ignore regulations which do not really apply and this could extend to a denial that there are, in fact, any ethical problems at all associated with the research. (139–40)

Those who argue for method- and discipline-specific ethical practice point to the particular historical contexts in which current ethical codes for research on human subjects were originally developed:

> Research practices have tended to generate controversy in proportion as the harms they pose to subjects are dramatic and dangerous. Perhaps as a matter of historical accident, it was experimentation in the biomedical sciences that first put subjects dramatically at risk of significant and irremediable harm. Not surprisingly, then, normative analyses of the ethical issues in research have along been marked by a preoccupation with cases drawn from biomedical experimentation involving human subjects (Beauchamp *et al,* 1982:4).

These authors argued that the dominance of the model of biomedical research in deliberations on research ethics partly reflects reactions to the abusive medical experimentation on non-consenting human subjects by the Nazis during World War II. Such experiments clearly contravened all four of the ethical principles outlined above, insofar as they refused to recognise the autonomy of the individuals concerned, inflicted harm which was vastly out of proportion to any benefit which was conferred (beneficence and maleficence) and ignored the principle of justice (see MacIntrye, 1982). In addition, as Beauchamp and co-workers (1982) pointed out, biomedicine can be seen as raising ethical issues in a particularly stark form, given its potential for putting people at dramatic and dangerous harm. While these authors accepted that biomedical and social science research have something in common, insofar as both involved the study of living persons, they argued that it is inappropriate to assume that ethical codes, developed in relation to biomedicine can simply be transferred to the social sciences, without careful consideration of the unique characteristics of social science research, which may indicate the need for some modification of the concepts or principles involved, in order to respect their spirit.

Cassell (1982) argued that discussions about ethics in research must take into account the way in which the relationship between the researcher and research participants varies between different kinds of research. She suggested that differences in the relationship between the researcher and the researched range along the following four dimensions.

- **The relative power of the investigator, as perceived by those being investigated**
  In biomedical research, the relationship between researcher and researched is asymmetrical, with the researcher being the more powerful partner in a hierarchical relationship. This is less likely to be the case in, for example, participant observation research,[129] where the researcher has relatively little power and participants are able to obstruct, refuse to cooperate and exert power over the researcher in a wide range of ways.

- **The relative magnitude of control over the context of the research setting**
  In biomedical experiments, control of the setting by the researcher is at the heart of experimental design. In qualitative studies, the power to control the setting is more likely to be shared between the researcher and the participants.

- **The relative magnitude of control over the context of the research**
  In experimental studies, the context is closely defined in advance of the research. The researcher is mandated to define what is relevant to the study in advance. The emergent nature of much qualitative research means that participants are able to influence the context of the research in ways which are not available to the subjects of experiments.

- **The relative magnitude of control over research interaction**
  Whereas, in experimental research, the researcher determines how much of the interaction between researcher and researched is relevant and how much should be treated as extraneous and potentially confounding variables, in qualitative research, "the paradigm is based upon human interaction in all its richness, variety and contradiction" (148).

Cassell (1982) concluded that the power relationships between researcher and researched are fundamentally different in different types of research and that this must be taken into account in applying ethical principles.

In the preceding discussion, we have linked social scientific and qualitative research together in discussions about the appropriateness of extrapolation from biomedical ethical codes. This reflects the discussion in the literature where such extrapolation is seen, by some at least as problematic for all social scientific research, whether qualitative or quantitative. Nevertheless, as Kelman (1982) pointed out, some methodological approaches are closer to biomedicine than others. For example, a participant observation study of a waiting room

[129] Participant observation is discussed in section 4.2.

in an outpatients clinic has much less in common with a biomedical experiment than, for example, a social scientific experiment in a psychology laboratory. It is therefore probable that, if there are any problems with the mechanical application of biomedical codes of ethics to social scientific research, these will be most clearly observable in relation to qualitative research. It is therefore not surprising that it is in this context that this issue has been most vigorously pursued (Cassell, 1978; Merriam, 1988).

In the remainder of section 4.7, we shall consider the arguments against the mechanical application of research codes derived from biomedical research to qualitative research under three headings:

- subject risks–potential benefits ratio (section 4.7.2)
- informed consent (section 4.7.3)
- confidentiality (section 4.7.4).

Throughout, we shall be examining not only the argument that ethical codes derived from biomedical research practice are inappropriately constraining for qualitative research, but also considering the possibility that such codes do not adequately take account of the risks which are specific to qualitative research.

4.7.2 Subject risks–potential benefits ratio

As discussed above, one of the ethical principles, which is routinely invoked in relation to research, is that of nonmaleficene. From this perspective, research is judged as ethical if it does no harm to those involved in it. More commonly, this principle of nonmaleficence is combined with beneficence, to argue that research is ethically acceptable if the benefits it offers outweigh its potential for harm. Codes of ethics and attempts at government regulation of research involving human subjects frequently require the researcher to demonstrate that the anticipated benefits of the research are likely to outweigh the harm likely to be caused to those who take part in it. A number of attempts have been made either to modify this risk–benefit calculation procedure, or to reject it as totally inappropriate, in relation to qualitative and/or social science research. The objections to the application of bio-medicine-derived risk–benefit analyses to qualitative research centre on three issues, which are discussed in the next three sub-sections.

4.7.2.1 Inappropriateness of affording the same degree of protection to all research participants

First, while the contention that researchers have a duty to avoid the introduction of unacceptable hazards to their subjects is generally accepted, there has been some debate about whether researchers owe the same degree of protection to all those they are studying. For example, Galliher (1980) has argued that not all research subjects require or deserve the same degree of protection:

> According to this reasoning, superordinates are not necessarily due the same degree and type of consideration by researchers as are other more deprived and powerless individuals. (Galliher, 1980:304.)

Galliher focused particularly on the rights of elites, including public officials, physicians and all those who might be said to be publicly accountable. He drew upon Nader's discussion of public accountability:

> Sociologists have the right (and perhaps also the obligation) to study publicly accountable behaviour. By publicly accountable behaviour we do not simply mean the behaviour of public officials (though there the case is clearest) but also the behaviour of any individual as he goes about performing public or secondary roles for which he is socially accountable – this would include businessmen, college teachers, physicians etc.; in short all people as they carry out jobs for which they are in some sense publicly accountable (Nader, 1969:365–6 cited in Galliher, 1980:298.)

Galliher argued that the social scientist does not owe the same duty of protection to such individuals as (s)he does to the less advantaged in society. Galliher (1973; 1980) suggested that superordinates are, by nature of their position, the repositories of power, which may be abused. Hence, the demands of public accountability mean that they are not owed the same duty of protection from harm as are the powerless in society.[130] Research which investigates the activities of such superordinates may well uncover practices that damage their reputation (or even employment prospects).

From Galliher's perspective, superordinates are in a much better position to protect their own interests than are subordinates and, hence, the researcher's responsibility to protect them is reduced. The implication of this position for research in HTA is that the researcher would not feel the same obligation to protect

[130] See section 5.3 for a discussion of even-handed treatment of all research participants.

administrators, physicians and other senior staff, from potentially adverse consequences of discrediting research findings as (s)he should in relation to, for example, patients. The particular types of harm to which such a position might leave a superordinate open are discussed in more detail below, but might include loss of self-esteem, stress, loss of respect from colleagues, loss of job security, legal jeopardy and the obligation to revise working practices. As Galliher acknowledged, the distinction between superordinates and subordinates may be difficult to operationalise in practice. Some people within an organisation (e.g. nurses or cleaners) may be subordinate to some groups, while wielding a considerable degree of power in relation to others.

If Galliher is right, and those who occupy a position of public trust do, in some measure, forfeit their rights to have their individual interests protected by the social scientific researcher, then this represents a difference between the biomedical researcher and the social scientist. This difference reflects not only the issue of public accountability, as discussed above, but also the assumption that superordinates are in a position to protect their own interests and do not require the protection of the researcher in quite the same way as the subject of a biomedical experiment would, whatever his or her status. However, we should note that removing the obligation to protect superordinates from harm arising from research does not necessarily involve their rejection of the principle that the risks and benefits of research should be weighed against one another. One could argue that, where superordinates are concerned, the potential harm of a given piece of research to powerful individuals is merely being offset against the potential benefit to the whole group of which such individuals are a part.

4.7.2.2 Relative harmlessness of qualitative research

The second objection to the application of risk–benefit analysis to qualitative and/or social scientific research is that, as the risk of harm to participants is so much less than in biomedical research, the application of such analyses is inappropriate. For example, Brewster Smith (1979) suggested that social scientists should focus on issues such as informed consent which emphasise research participants' autonomy, rather than risk–benefit analyses. Some, such as Patullo (1982), Diener and Crandall (1978), Cassell (1978), Brewster Smith (1979) and Cassell (1982) have argued that social research is relatively harmless, particularly when compared with biomedical research. While the latter may lead to severe physical harm or even death, the former is seen as carrying few risks.

Cassell (1982) suggested that different kinds of research can be placed on a continuum in terms of the risk of harm which they pose to those who take part in them. Biomedical and psychological experiments are seen as having the greatest potential for harm, with survey research having slightly less and qualitative research posing minimal risks: "In the conduct of fieldwork, there is a comparatively minimal level of harm – again, primarily violation of anonymity or confidentiality" (Cassell, 1982:149).

Cassell does not go on to argue that this means that qualitative research should not be open to ethical scrutiny. Rather, she argued that by focussing upon a risk–benefit calculation, the researcher is in danger of obscuring the very real ethical difficulties which do arise in qualitative research. She suggested that it is more appropriate to consider the ethical aspects of qualitative research in terms of the principle of autonomy.[131]

Most authors agree with Beauchamp and coworkers (1982) that the risks of harm to the subjects of social scientific research are rarely of the same order as the dramatic risks of serious and possibly irremediable physical harm characteristic of some forms of biomedical research. However, a number have challenged the assumption that such risks are therefore negligible, pointing to the considerable potential which social science research, and by extension qualitative research, has for causing harm not only to the individual (in terms of not only physical but more pertinently psychological harm, damage to self esteem and/or to interpersonal relationships) but also to society as a whole or sub-groups within it (through the erosion of trust, victimisation, blaming the victim, stereotyping, etc.) and to the research profession (as a result of public hostility towards research arising from outrage at the behaviour of some researchers). Warwick (1982b) outlined the range of risks which social scientific research may pose to its subjects. Of those which he included in his list, two have particular relevance to qualitative research.

1. **Risk of psychological abuse or injury**
Such risks might include anxiety, stress, guilt and damage to self-esteem. These risks may

[131] See section 4.7.3 for a discussion of autonomy.

arise both during the process of data collection and as a result of the publication of data (Cassell, 1979). Merriam (1988) raised the possibility that in-depth interviewing may have unanticipated long-term effects upon those who take part:

> What are the residual effects of an interview with a teacher who articulates, for the first time perhaps, anger and frustration with his position? Or the administrator who becomes aware of her own lack of career options through participation in a study of such? Or the adult student who is asked to give reasons for failing to read? (180)

Or, one might add, the patient who, in the course of an in-depth interview, articulates for the first time the risks associated with their medical condition, or the frustration experienced in interactions with health professionals, or comes to realise the part that their own behaviour has played in the causation of a disabling or life-threatening disease. Of course, as Merriam acknowledged, such discoveries may be positively beneficial, but they are not necessarily so. The problems which the indeterminacy of harms and benefits in social science research create for harm–benefit analyses are discussed below.

Kelman (1982) also suggested that interview informants may experience stress or discomfort during an interview, either because they are embarrassed by their lack of information on the topic on which they are being questioned, or because they have no opinions on topics on which they feel that they are being expected to have opinions. On the other hand, they may feel uncomfortable or embarrassed about the opinions they do hold.

Research participants may also be put at emotional and psychological risk in participant observation research.[132] Ironically, perhaps, this risk may be greatest where the participant observer has formed warm and respectful relationships with those who are being studied. While holding that "the risk from interaction during fieldwork is rarely as great as that from medical or other behavioural experimentation", Cassell (1978; 1979) acknowledged that: "There is an emotional risk posed by the observer who takes pains to become part of a group's ongoing life and be defined as **friend** and on occasion advocate and who then leaves, breaking off ties with those who were studied" (Cassell, 1978:138 original emphasis).

These risks may be compounded where such observational research has been covert rather than overt. The ethical problems raised by such covert research are discussed more fully below.[133] However, the risks of covert research to the psychological and emotional well-being of those who subsequently discover that a person, who was accepted into a group or network as a member, or even friend, was in fact pursuing his or her own research agenda may be considerable.

2. **Risk of damage to the individual or group's interests**

Cassell (1978, 1979) argued that, unlike experimental research where the greatest risks arise during the period of experimental manipulation and data collection, in qualitative research, it is the publication of research data which poses the greatest threat to research participants (see also Wax and Cassell, 1979). Research findings are not morally neutral and, once they are in the public domain, the researcher retains little control over the uses to which they are put. While guarantees of confidentiality[134] may help to protect individuals from the most negative consequences of publication, they will not necessarily prevent damage to the interests of the group of which research participants are members. Particularly where research is carried out on the relatively powerless or disadvantaged, the findings may be used to either control or manipulate those who have been studied (Cassell, 1979; Kelman, 1982; Finch, 1984).

Burgess (1985) described some of the ethical dilemmas, which arose in relation to the publication of data from his participant observation study in a comprehensive school. In the course of this study, he studied the Newsom Department, a unit for low achieving pupils within a comprehensive school. He observed a range of teacher and pupil behaviour within the Newsom Department that would not have been tolerated within the rest of the school. He described the reaction to an oral presentation of his data among the staff at the school:

> My presentation focused on life in the Newsom classes and I took extracts from my field notes and records from teachers' diaries to use as illustrations of the activities that occurred in Newsom classes. The picture I depicted included a situation where a Newsom pupil had sat in a classroom and proceeded to dye his hair in the middle of the lesson. On hearing this the headmaster remarked, 'I find it impossible to believe

[132] See section 4.2 for a discussion of participant observation; [133] See section 4.7.3.5; [134] See section 4.7.4.

> that any member of this staff would allow this to happen in a lesson'. Several teachers disputed the head's remark and quoted instances of other activities within classrooms. However, it was evident that they carefully avoided any reference to their own teaching in MacGregor School, and to work by other teachers within the school. ... Meanwhile the teacher in whose classroom this incident had occurred looked distinctly uncomfortable throughout the discussion. (155)

Burgess did not identify the teacher concerned, who thanked him afterward, saying that if he had done so this would have jeopardised the teacher's chances of promotion. However, Burgess recognised that, particularly in a participant observation study, where the setting studied may be relatively small, it is not always practical to protect individuals within the setting:

> In presenting my data I had disguised the names of all the teachers and pupils in my study, but because there were only four members of the Newsom Department, it was possible for the head and for other teachers in the school to have a reasonable chance of identifying them. Indeed, on occasions, I have found that the head has called individuals by name other than the pseudonym when discussing my study. In some cases he has guessed correctly, while in other he has been incorrect. However, I have never confirmed or denied any of the guesses that have been made as I would hope that this would result in my data being used neither for nor against my former colleagues. (156)

In this situation, the individual teachers were put at risk of harm, to their careers as well as to their self-esteem, by the publication of data. This points to the way in which, particularly in participant observation studies, while it may be possible to protect the anonymity of the setting studied, it is less possible to protect individuals from others within that setting. The same problem could well arise in similar studies in medical settings.

As Burgess (1985) pointed out, even if it had been possible to protect the individual teachers within Newsom, this would not have resolved all of the ethical issues raised by his study. Indeed the headmaster of the school challenged him for having failed to protect the interests of the children in the Newsom Department by reporting the kinds of activities that took place there. The headmaster argued that, as a result of his failure to do so, several generations of children had passed through a department where, he observed, they received "what could hardly be described as an education".

Here, Burgess was confronted by a conflict between his loyalty as a researcher to those studied, and an alleged wider responsibility to reveal practices that are deemed to be against the interests of a vulnerable group within the setting. Clearly similar dilemmas are likely to arise in participant observation studies carried out in healthcare settings. The researcher may find him or herself in possession of information about unethical or rule-breaking behaviour within such settings, which would harm the interests of those studied, if revealed. Here the principle of beneficence can be seen to be in conflict with that of non-maleficence.[135] As Burgess argued, there is no slick solution to these dilemmas, and indeed, in many cases researchers may be obliged to seek the least bad compromise.

Burgess also raised another aspect of the dilemma which may be of relevance to research in healthcare settings. Burgess noted that his detailed study of the behaviour of Newsom Department and their teachers may well be used to control such behaviour in the future. From the headmaster's perspective, this was clearly seen as one of the potential benefits of the study. However, as Burgess pointed out, if the *status quo* in the Newsom Department reflected an attempt to accommodate a situation in which de-motivated pupils are obliged to stay on at school, with no hope of benefit, undermining this may in fact be interpreted as harming the pupils' interests:

> I am aware that my material can be read with a view to considering how pupil behaviour can be controlled and more 'education' provided. In such situations the teachers' activities may be out of professional concern for the 'education' of pupils, but in turn this may result in some curtailment of the strategies that pupils adopt to come to terms with their schooling. (157 original emphasis)

In healthcare settings, as elsewhere, knowledge represents power and such knowledge may be used to manipulate or control those the relatively powerless. Burgess argued that researchers must grapple with the unintended as well as the intended consequences of their research.

Finch (1984) argued that the very nature of the in-depth interviewing,[136] especially where both

[135] See beginning of section 4.7. This is likely to be a particular problem for those researchers who are also members of the professional community under study. There is a potential conflict between research and professional ethics;
[136] In-depth interviewing is considered in section 4.3.

interviewer and interviewee are women, means that women may be lulled into revealing information which may not, in the final analysis, be in their own interests to disclose. Writing from a feminist perspective, Finch criticised the argument that there is a tension between the short-term desire of citizens to protect their own interests and the longer-term goal of sustaining informed criticism in a democratic society. She argued that in a patriarchal society, the longer-term interests, which are used to justify the sacrifice of women's short-term interests are unlikely to include the particular interests of women. She emphasised that, in research on women, it is not only individual but also collective interests that are at stake. She concluded that:

> The latter [individual interests] may be relatively easily secured with guarantees of confidentiality, anonymity, codes of ethics and so on. It is far more difficult to devise ways of ensuring that information given so readily in interviews will not be used ultimately against the collective interests of women. (83)

4.7.2.3 Practical difficulties in application of risk–benefit analyses to qualitative research

The third objection to risk–benefit analysis is that it is simply impractical in relation to much social scientific research. The very possibility of implementing the kind of risk–benefit analysis, which is routinely used in assessing biomedical research, has been questioned by some authors (Beauchamp *et al,* 1982). Some (e.g. Cassell, 1982) have argued that advance assessment of harms and benefits raises considerable practical difficulties in social science research generally, and qualitative research in particular. Brewster Smith (1971) recorded how the group charged with preparing the American Psychological Association's code of ethics: "... discovered early on, there is no way at all to conceive of a workable cost/benefit formula that could resolve actual issues of research ethics in the social sciences" (16).

From Brewster Smith's perspective this does not imply that no attempt should be made to consider studies in the light of their likely harms and benefits, but he argued that such a framework cannot be used to calculate whether or not a given study is ethical.

A particular problem facing social, as opposed to biomedical scientists, is that, while the costs of social research are most likely to be incurred by the individuals studied, the benefits are likely to be to science and to society at large. As Brewster Smith argued, this is very different to much clinical research, where it is often possible to weigh the risks to the patient of a particular treatment, against the possible benefit to that patient.

MacIntyre (1982) emphasised the differences between qualitative and experimental research in terms of the degree of predictability of potential harms and benefits:

> The paradigm case for an effective use of cost–benefit analysis in social science research is the well-designed controlled experiment, conducted with a view to an immediate decision between alternative policies. If the experiment is well-designed, then the population affected by the experiment, whether in the experimental group or the control group, will be clearly identifiable. If any harm is done to them, we shall surely be able to say with a useful degree of clarity **who** paid certain costs and **what** costs they paid. If the experiment provides a reasonably adequate test for deciding between alternative policies then we shall surely be able to say with a useful degree of clarity **who** will receive certain benefits and **what** benefits they will receive ... it is only an effectively usable instrument where we are dealing with the more rather than the less predictable, with the more rather than the less quantifiable, with the more clearly rather than the less clearly definable. It is for this reason that various types of experimental and quasi-experimental research, various types of survey research, and various types of fieldwork can be placed on a spectrum such that cost–benefit analysis becomes, as we move across that spectrum, a progressively less useful method of evaluation. Note that this is **not** because there are not both costs and benefits at the fieldwork end of the spectrum, but because if they are to be mobilised effectively in an overall evaluative argument, it will have to be in some way other than within the framework of cost–benefit analysis. (180–1 original emphasis)

The difficulty of predicting the positive and negative consequences of participation in qualitative research studies relates to the argument that the greatest risks from experimental research occurs during the research study itself, whereas the greatest risk in qualitative research arises in relation to the dissemination of research findings.[137]

A second practical problem associated with the application of risk–benefit analysis to social scientific research lies in deciding what constitutes a benefit and what constitutes harm. Unlike biomedical research, where there may be substantial consensus about what constitutes a harm, the same is not necessarily true of social scientific research.

[137] See section 4.7.2.2.

As Brewster Smith (1979) has argued, we cannot assume that an increase in self-knowledge invariably represents a benefit. Likewise, as Burgess's discussion of the ethical issues raised in his study of Bishop MacGregor's School demonstrated,[138] not all groups affected by research will benefit or be harmed by research in equal measure.

4.7.3 Informed consent

MacIntyre (1982) argued that, even in the extreme case of the Nazi experimentation on living persons, the fundamental ethical objection was not primarily to the harm that was caused to the experimental subjects, but rather to the failure to obtain their informed consent. In MacIntyre's terms these subjects were wronged rather than harmed. This is a significant difference, since MacIntyre argued that wrongs cannot be incorporated into risk benefit analyses. Whereas harms and benefits can be weighed against one another, wrongs and benefits are incommensurable.

Beauchamp and co-workers (1982) suggested that the prominence of the issue of informed consent in debates about research ethics reflects its significance in relation to legal judgements and malpractice suits, particularly in the field of biomedicine. They suggested that this may be another example of the imbalance which may occur when ethical principles derived from biomedical research are mechanically applied to social science research. The implication of their position is that one must consider in detail the ways in which the application of the principle of informed consent may enhance or indeed undermine the autonomy of participants in qualitative, as opposed to experimental research.

A number of writers have questioned the appropriateness of a rigid application of the principle of informed consent to social science research in general and to qualitative research in particular. For example, Murray Wax (1977; 1980) argued that the differences between qualitative research and biomedical research are so great as to render the formulation of the principle of informed consent, which was developed in relation to the latter, inappropriate to the former: "Since fieldwork involves such a different methodology, with unique and often unpredictable benefits and risks (for both researcher and researched), this regulatory system is inappropriate in many ways, particularly on the issue of consent to the ethical dilemmas of fieldwork" (Wax, 1977:29 cited in Macklin, 1982.).

Trend (1979) identifies three components of informed consent which may be identified in codes of ethics and other regulations. These are:

- that informed consent should be voluntary
- that informed consent should be knowledgeable
- that those giving informed consent should be competent to do so.

We shall consider the application of informed consent to qualitative research under these three headings.

4.7.3.1 Informed consent should be voluntary

The emphasis upon informed consent, particularly within biomedical experimentation, is an acknowledgement of the power which the experimenter, particularly the biomedical experimenter, has to coerce or unduly influence potential research subjects to participate in research. The ethical requirement is that consent from research participants should be given voluntarily. Cassell (1979) argued that the relationship between the researcher and the researched is very different in fieldwork from that which exists in biomedical or psychological experimentation.[139] In the latter, the experimenter is in a powerful position *vis à vis* the research subject. This greater degree of power is intrinsic to experimental design where, as Cassell said, the researcher acts and the subject reacts. By contrast, this power differential is not intrinsic to fieldwork, where the researcher and research participant act and interact on each other.

Cassell (1979) conceptualised informed consent as an attempt to equalise the unequal power relationship between researcher and research subject. As such it is more appropriate to the experimental situation, where power differentials are greatest, than to qualitative research, where control over what happens is shared between researchers and participants.

Turning to the requirement that informed consent should be knowledgeable, we encounter a further set of difficulties in applying this principle to qualitative, rather than experimental research. As with risk–benefit analyses, doubts have been raised about the extent to which fully informed consent, in advance of the study, is feasible in much qualitative research. In particular, questions have been raised about what 'fully informed' can mean in a situation where the research design is emergent during the course of the research. Cassell (1979)

[138] See section 4.7.2.2 for details of this study; [139] See section 4.7.1.

has argued that informed consent is more feasible when it is possible to spell out what will happen during the process of research: "Because of the limited, controlled, and unidirectional nature of the relationship between experimenter and subject in traditional experimentation, it is comparatively easy to foresee the risks a subject will be exposed to, and consequently, to describe them in a consent form" (139).

This situation rarely applies in qualitative research, where, through progressive focussing and theoretical sampling[140] the details of the research study are negotiated during the study itself (Hammersley and Atkinson, 1995). For this reason, fully informed consent, in advance of the fieldwork is rarely possible. Rather, qualitative researchers are more likely to negotiate initial entry to a setting, and then engage in ongoing negotiations about what they will observe, who they will talk to etc., over a period of time.

Murray Wax (1980) fully acknowledged the importance of informed consent in experimental research. He defined the relationship between experimenter and subject as a contractual one. In such research: "When consent is solicited, the subjects are treated as autonomous beings, aluable and competent in their own right, and the scientist is freed of any taint of authoritarian or coercive conduct" (284).

Relationships in qualitative research, on the other hand, are more likely to be intimate rather than contractual and this has consequences for the way in which consent should be obtained.

These writers have not claimed that the researcher is excused from the demand that research should only be carried out upon consenting individuals. Rather, they suggested that the signed consent form, typical of biomedical and psychological experimental practice, may be an inadequate protection of the rights of qualitative research participants. As Murray Wax (1980) put it:

> Under these circumstances, **consent** becomes a negotiated and a lengthy process – of mutual learning and reciprocal exchanges – rather than a once-for-all-event. ... Needless to say, the conventional 'consent form' is so irrelevant as to be a nuisance to all parties. ... As hosts, their [research participants'] judgement is ... sequential and conditional so that consent is a continual process, dependent on mutual learning and development. (275 original emphasis)

Truly informed consent, in the context of much qualitative research, is not obtained by offering participants the option of refusing to take part at the outset. Rather it demands from researchers a commitment to the autonomy of the research participants and a willingness to withdraw or to amend their emergent design, in the light of participants' autonomous wishes, at any stage of the research.

The principle of fully informed consent requires not only that the researcher should describe the intended research, but also that (s)he should ensure that potential participants are fully aware of any risks arising from the study. In qualitative research, as discussed above, such risks may arise not only in the process of the study (as in experimental interventions) but also at the later stage when data are disseminated. As Dingwall (1980) argued, ensuring that consent is fully informed in this way is not straightforward: "It seems likely that many participants in sociological research are oblivious to the promptings of the most conscientious investigator, especially in areas of their life with less immediate and acute personal consequences" (877).

Dingwall compared this situation to the difficulty that gynaecologists report in persuading patients who seek tubal ligation to consider the long-term consequences of such a decision, with the result that a significant proportion seek reversal of the operation in due course. He argued that even where the researcher is fully committed to obtaining consent from participants, it may be difficult to persuade such participants to engage with discussions about the possible risks of research which may, both from their perspective, and sometimes in reality, be remote.

4.7.3.2 Competence to give informed consent

The third requirement is that fully informed consent should be given by those who are competent to do so. This raises another difficulty in relation to qualitative research. As Brewster Smith (1979) wryly suggested, it may be difficult to fully explain the purposes of research "without sending informants and cohabitants to graduate school" (14).

One illustration of the way in which fully informed consent may be compromised by participants' lack of understanding of the theoretical under-pinnings of research can be found in the methodological appendix to Buckholdt and Gubrium's participant

[140] See section 4.1 for a discussion of theoretical sampling. Progressive focussing is discussed in section 4.7.1.1.

observational study of a treatment facility for emotionally disturbed children (Buckholdt and Gubrium, 1979).[141] The authors described their painstaking negotiation of consent from all staff members in the facility. They met formally with the facility director, and also with each separate professional group within the facility, to explain what would be involved in the study. This was augmented by a written description of the study. They explained that they would be studying various professional practices throughout the facility. However, their understanding of the nature of practice was grounded in their own theoretical position and differed markedly from the way in which this term was interpreted by research participants:

> They understood practice to be a matter of application and our study as being done in the service of improving the application of their skills to the treatment, care and evaluation of emotionally disturbed children. Our own understanding of practice, on the other hand, was more basic in that we were, in effect, interested in the practice of professional applications of skills. (253)

As well as having difficulty in understanding the purpose of the proposed research, participants may also be hampered by unfamiliarity with qualitative methods. As Finch (1986) has suggested, qualitative research methods are less well understood by the general public than are methods such as surveys and experiments.

4.7.3.4 Constituencies from which informed consent should be sought

In addition to the three aspects of informed consent identified by Trend (1979), a fourth deserves consideration. Here we are concerned with the constituency from which informed consent should be sought. This concern relates again to the differences between much experimental and qualitative research. Experimental research, particularly in biomedicine and psychology, involves experimental treatment of individuals. In this case it is relatively clear who is at risk from the experimental treatment and therefore whose consent is required. Qualitative research, on the other hand, often focuses upon groups rather than individuals, upon social settings rather than upon research subjects. This immediately raises the question of whose consent should be sought before research may proceed. Here the distinction between pragmatic concerns about securing access and ethical concerns about individuals' rights to self-determination are highlighted. As Dingwall (1980) pointed out, in any stratified setting the consent of those at the top of the hierarchy is likely to be needed if a study is to go ahead.[142] However, if one's concern is with autonomy, rather than pragmatics, one is obliged to consider whether all those who are involved in the setting to be studied are not owed the same right to give or withhold their consent.

In practice, many qualitative researchers do not offer all those involved in a setting the opportunity to withhold their consent. For example, Burgess (1985) described how, while his research role was known to the staff in the school which he studied from the outset, pupils were neither informed nor consulted about the research study, at the outset. Similar situations might arise in studies of health-care settings if, for example, senior administrators, nurses and/or physicians were consulted, but cleaners or other ward staff were not.

Dingwall (1980) also pointed out that, even where there is a commitment to obtaining informed consent from all groups within a setting, serious practical difficulties may remain. Given that the researcher is typically involved in the setting for an extended period of time and that, unlike in experimental research, no attempt is made to control access to that setting, the process of obtaining informed consent from each and every person in a setting may prove unwieldy:

> So many people are encountered casually that it is impractical to obtain consent on each and every occasion without causing total disruption. ... In any personal service organisation, one is likely to encounter a wide range of professional and ancillary staff who may not actually form part of the study but may be a source of relevant data on the particular group under scrutiny. (878)

This difficulty in obtaining informed consent from all participants in a setting raises the further question about whether a researcher has an obligation to obtain informed consent when the study is being conducted in a public setting. Murray Wax (1980) suggested that, where the researcher is studying behaviour which is routinely publicly observable, then the requirement of informed consent appears dubious. He argued that it is not reasonable to ask field workers to obtain informed consent when their research involves doing no more than any member of the public is free to do.

[141] See appendix 1 for details of this study; [142] Covert research is considered in section 4.7.3.5.

4.7.3.5 Covert research

Covert research can be seen as a particular case in which the principle of informed consent is called into question. The covert researcher does not simply neglect to inform participants in a setting of his research intentions (either fully or partially, as discussed section 4.7.3.4). Rather (s)he chooses to deliberately mislead the participants in the setting.

The ethical debate in relation to qualitative research has been dominated (some would argue inappropriately) by the controversy about whether or not covert research methods are ever justified. Much of the discussion has centred upon two famous, or by some accounts infamous, studies. The first was Humphrey's study of casual homosexual encounters in public lavatories in the USA, published under the title *Tearoom Trade* (Humphreys, 1970). Humphreys carried out covert observations in public lavatories to collect the data for his study. He played the established role of 'watch queen' or lookout for men engaged in homosexual acts. He also identified the men he had observed, using the registration numbers of their cars, and subsequently carried out research interviews with them, in their own homes, posing as an interviewer in an anonymous health survey.

A second covert observational study, this time in the UK, which generated heated controversy, was Homan's study of an old-time Pentecostal assembly. Homan 'joined' the Pentecostal assembly that was the focus of his study (Homan, 1980). He described the role he played:

> The field observer conformed his outward behaviour in all possible respects with the norm existing in the assembly. He adopted the appropriate postures for prayer, singing and listening; in singing he allowed his voice to be audible; in listening to addresses and announcements he interpolated the 'praise phrases' as appropriate. He carried a black leather bible with him to the assembly. He shook hands with other members of the assembly and exchanged sacred greetings, thereby presenting himself as a 'saint' rather than 'sinner' and pre-empting the special attention (evangelism) given to outsiders. He took standard initiatives like interrupting a hymn and reciting the forthcoming verse: he occasionally requested choruses. (49 original emphasis)

While these two studies represent a fairly extreme form of covert observation, they raise ethical issues, which apply to many observational studies (see also Festinger *et al,* 1956; Lofland and Lejeune, 1960; Rosenhan, 1973). Much of the debate has revolved around the issue of whether ends may indeed justify means. In defending their research practice, those, such as Humphreys (1970) and Homan (1980), who have carried out covert research, have pointed either to the positive benefits (Humphreys) or the lack of harm (Homan) which has accrued to the group studied, as a result of the research study. Critics, such as Dingwall (1980), Bulmer (1980) and Beauchamp and co-workers (1982) have not only dispute the claim that such research causes no harm, to the individual, the group to which (s)he belongs, or society as a whole, but have also argued that the consequences of such research are not the only criterion by which it is to be judged.[143]

Beauchamp and co-workers (1982) argued that covert research is morally controversial because such deception may be seen as being disrespectful to those studied, in ways which contravene the principle of autonomy, irrespective of whether the consequences of such research are, on balance, beneficial or harmful. In other words, the appropriateness of using a harm–benefit analysis is disputed. It is argued that the ethics of covert methods should be discussed in terms of the rights rather than consequences. However, as Beauchamp and co-workers (1982) pointed out, conducting the discussion of the ethical dimensions of covert research in terms of rights rather than consequences can raise further controversy. They cited Horowitz and Rainwater's defence of covert methods in terms of the researcher's right to pursue and communicate knowledge and their resistance to any attempt to curtain covert research that could be seen as interfering with the social scientist's duty to extend knowledge.

Defenders of covert studies draw upon a number of lines of argument. Some, such as Homan (1980) have argued that covert research is sometimes in the best interests of those being observed. He claimed that, in his own study, covert methods may:

> ...constitute a more considerate and sensitive strategy to those in which conspicuous recording devices and modes of behaviour or the self-disclosure of the researcher cause the performers to know that their activities are being monitored. Such a self-consciousness as a disclosure of the researcher may engender would adversely affect the character and quality of the performance for the subjects themselves: for example it may become less sincere. (53)

[143] See beginning of section 4.7 for a discussion of deontological and consequentialist ethics.

The flaw in Homan's argument is that he assumed that, with or without informants' consent, the research will go ahead. He arrogated to himself the right to decide what is in the best interests of the participants. In doing so, he not only rejected the principle of participant autonomy but also advocated what were, in essence, coercive tactics.

The second argument which is advanced by advocates of covert research is a methodological one. This is proposed not only by qualitative researchers such as Homan, but also by advocates of deceit in psychological experiments, such as Milgram (1963). It is held by these authors that, where disclosure of the nature of the research, or indeed that research is taking place at all, would jeopardise the validity of the research, then disguised or covert research is justified. In relation to qualitative, observational studies, the argument is advanced that the overt presence of a researcher is likely to distort the behaviour of participants in a setting to such an extent that it may render the findings invalid.

A number of authors (Bulmer, 1980; Dingwall, 1980; Wax, 1980) have challenged this defence of covert methods. Dingwall (1980), for example, rejected this argument on the grounds that it exaggerates the difficulty of obtaining valid data using overt methods. He argued that, even where participants in a setting are aware of a researcher's presence and purpose, the demands of 'task performance' are likely to be a more important constraint upon participants, especially once the novelty of the researcher's presence has worn off. Similarly, Bulmer (1980) suggested that covert methods are often simply unnecessary, since the same objectives can be achieved by open methods. It is empirically the case that researchers have been granted access to highly sensitive settings, such as the civil service in Britain (Heldo and Wildavsky, 1974), the Mafia (Ianni and Ianni, 1972), the activities of professional 'fences' (Klockars, 1977), professional criminals (Polsky, 1971), drug barons (Adler, 1985) and the National Front (Fielding, 1982).

Critics of covert research (e.g. Erikson, 1967; Bulmer, 1980; Dingwall, 1980) have raised a number of objections to the practice, some of which concern issues of autonomy[144] and some that are concerned with the possible harm[145] which such studies may cause to the people studied.

Bulmer (1980) argued that covert research involves betrayal of trust and that this breach is not necessarily mitigated if the researcher is successful in anonymising the research setting:

> The preservation of anonymity and confidentiality does not preserve them from harm. If those studied subsequently read or learn of the publication of the research, they must come to terms with the fact that they have been cheated and misled by someone in whom they reposed trust and confidence. (61)

Similarly, Erikson (1967) argued that sheer act of entering a human transaction on the basic of deliberate fraud may be painful to the people studied. In addition, covert research may also represent a gross invasion of personal privacy (Bulmer, 1980) and an exhibition of tacit disrespect for those studied.

Alongside these risks to the individuals who are the objects of research, a number of authors (Erikson, 1967; Bulmer, 1980; Dingwall, 1980; Kelman, 1982) have pointed to the wider implications of covert research. On the one hand, it is feared that such research will contribute to the erosion of trust in society as a whole (Kelman, 1982; Warwick, 1982a). Warwick argued that covert research: "...ultimately helps produce a society of cynics, liars and manipulators and undermines the trust that is essential to a just social order" (58).

On the other hand , the use of covert research methods may undermine future research by social scientists (Erikson, 1967; Bulmer, 1980; Dingwall, 1980).

The current consensus appears to be that covert research is rarely, if ever, justified. However, a number of authors have pointed out that creating a dichotomy between covert and overt research may be misleading (Roth, 1970). Roth argued that all research is secret in some ways and to some degrees, as researchers never tell their subjects everything. The difficulties of obtaining fully informed consent from all possible participants in a setting were discussed in section 4.7.3.4. Similarly, as Hammersley and Atkinson (1995) pointed out, even where the researcher's purpose has been made explicit, it is not uncommon for participants in a setting to forget the ethnographer's research role, once they come to know him or her as a person. However, it may be important, in this context, to recognise the distinction between a researcher who deliberately sets out to conceal his

[144] See section 4.7.3; [145] See section 4.7.2.

identity and/or purpose from those he studies, and one who finds it impossible to ensure that all those (s)he studies are fully informed about all aspects of the research programme. It is the former, rather than the latter, which is most likely to constitute a betrayal of trust.

4.7.4 Confidentiality

There is a consensus that researchers have an obligation to do their utmost to protect participants anonymity and keep data in confidence unless specific agreements have been made to the contrary (Bulmer, 1982; Punch, 1994). As Bulmer (1982) put it: "Identities, locations of individuals and places are concealed in published results, data collected are held in anonymised form and all data are kept securely confidential" (225).

A number of the issues surrounding confidentiality, which are specific to, or magnified in, qualitative research, have already been discussed in section 4.7.2 in relation to the risk that research participants may be harmed as a result of their involvement in qualitative research. In this section, we shall briefly discuss the aspects of qualitative research that may make protecting participants' anonymity particularly difficult.

Beauchamp and co-workers (1982) argued that, while questions of privacy and confidentiality may arise in certain kinds of biomedical research, they assume a more central position in discussions of the ethics of social science research. In particular, qualitative research is seen by some as posing particular difficulties in this respect (Finch, 1986). In quantitative research, confidentiality of data may be treated as a purely technical matter, to be managed through rigorous procedures for data handling and anonymisation. Finch argued that: "Questions of confidentiality are magnified in qualitative research because of the position of trust which the researcher aims to develop, and because research on a small scale makes individuals much more easily identifiable" (203).

On the one hand, as a result of the research methods used, the qualitative researcher may come into possession of more sensitive information about participants than is routinely the case in more quantitative research. On the other, the nature of qualitative research, and in particular participant observation,[146] may make it considerably more difficult for the researcher to ensure that such data, once published, is totally unattributable. Whereas, for example, the recorded replies to survey questions, are likely to contain little in the way of extraneous material, which might help to identify respondents, in recordings of qualitative interviews, or in field notes from observational studies, there is likely to be a much greater level of detail and therefore, in their raw state, these may well contain information which would permit identification of research participants.

Given the greater intimacy of the context in which qualitative data is often collected, it is also more likely that such data will contain highly intimate material, which might embarrass or even compromise the research participant, if it were to be published in an attributable form. Similarly, as Kelman (1982) remarked, participant observers may be: "...privy to personal information about individuals and to sensitive information about groups, organizations and communities, disclosure of which might cause them embarrassment and damage their material interests in a variety of ways" (84).

As Munhall (1993a) pointed out, both the informality and the long-term nature of researcher–researched relationships in qualitative research may raise particular problems in relation to confidentiality. Reiss (1979) has argued that: "The single most likely source of harm in social science inquiry is that disclosure of private knowledge can be damaging" (73).

The principle of informed consent[147] is intended, among other things, to ensure that research participants are in a position to protect their interests against such disclosure, if they so choose. In participant observation studies,[148] and even in qualitative interviews,[149] it is not uncommon for the participants, who at one level may be fully aware of the researcher's purposes, to become so familiar with the researchers' presence or so preoccupied with what Dingwall (1980) has described as **task performance**, that they drop their guard and treat the researcher as a friend or intimate, exposing themselves to the risk of disclosure in the process. Similarly, in participant observation studies, it may not always be clear to participants when the researcher is on or off duty. Typically, participant observers treat all of their contacts with participants as a legitimate source of data. Dingwall (1980) described how this led to a misunderstanding in his study of health visitor training:

[146] See section 4.2; [147] See section 4.7.3; [148] See section 4.2; [149] See section 4.3.

> During my research on health visitor training, I ran into a difficult patch after about six weeks when the students realised that I was recording backstage events and conversations. They found it difficult to link this to the announced theme of studying health visitor training although I saw it as a logical development for the study of professional socialization in attempting to locate that experience more firmly in the competing pressures on students' time and in their overall life-trajectory. (879)

Whether the participants have become so familiar with the researcher's presence that they are effectively unaware of the research dimension, or they perceive the researcher to be off-duty, they may act or speak in ways which are not intended to be on the record. In some ways, such a degree of acceptance within a setting may be seen as a methodological success, insofar as it reduces the possibility that research data is merely an artefact of the researcher's presence. On the other, it poses a heightened set of ethical dilemmas for the researcher who must confront the possibility that any breach of confidentiality would represent not only a violation of the research relationship, but also of personal trust. As Bulmer (1982) pointed out, the publication of such data may leave research participants feeling that they "have been cheated and misled by someone in whom they reposed trust and confidence" (15). Such data may also include material that puts the research participant in legal or professional jeopardy. Finch (1984) also argued that, particularly women informants, may disclose information, in the context of qualitative research interviews, which leave them vulnerable.

The greater level of personalised detail in qualitative research materials may also such data attractive to official groups, including the police and the courts. Finch (1986) reminded us that researchers may be vulnerable to legal obligations to hand their data over in this way. For example, Trend (1979), described an attempt by the General Accounting Office, a US government agency which monitors expenditure of public funds, to audit data which had been obtained, under a written guarantee of confidentiality, from low income households.

All of this means that participants in qualitative research may be at particular risk, if the confidentiality of the data is compromised. As Kelman (1982) suggested, when it comes to individual data, careful procedures can offer some protection. Such procedures might include removing identifying information from data at the earliest opportunity, the use of pseudonyms for individuals and other aspects of the setting under study, the alteration of non-relevant details of the setting for the purposes of camouflage, and so on (Burgess, 1985). However, as Burgess pointed out, such measures are not always successful and readers of research reports may believe, sometimes wrongly, that they have identified the setting under study. Burgess's own practice was never to confirm or deny such speculation.

Punch (1994) described some of the ways in which the confidentiality of data may be breached:

> The major safeguard to place against the invasion of privacy is the assurance of confidentiality. But even such assurances are not watertight, and 'sociologists themselves have often flagrantly betrayed confidence, undoing all the work of covers, pseudonyms, and deletions' (Rock, 1979). I mentioned earlier the tendency to choose sites close to one's university; pseudonyms can often be punctured by looking up the researcher's institutional affiliation at the time of the project ... Holdaway (1982) painstakingly uses a pseudonym for his research police station, but then refers in his bibliography to publications that make it plain that he studied the Metropolitan Police of London. ... In addition, the cloak of anonymity may not work with insiders who can easily locate the individuals concerned or, what is even worse, **claim** that they can recognise them even when they are wrong. Many institutions and public figures are almost impossible to disguise, and, if they cooperate in research, may have to accept a considerable measure of exposure, particularly if the popular media pick up on the research. (92 original emphasis)[150]

4.7.5 Summary

In this section we have considered some of the ethical issues which arise particularly in relation to the conduct of qualitative research. While, at one level, these issues may be considered to be less pressing than in much biomedical research, since the risk of serious and irremediable harm is usually less, their complexity and unpredictability raises particular problems. We have argued that research participants are at risk of significant harm in many qualitative research studies, either to their psychological or emotional well-being or to their professional or even legal position. Whereas, in biomedical research such harm is most likely to

[150] Social science researchers are currently under considerable pressure to disseminate their findings beyond the academic community. In doing so, they increase the risk that their research will be taken up in ways over which they have minimal control or even influence.

occur during the research study itself, in qualitative research the greatest risk may arise as the results are disseminated. Qualitative research also raises particular difficulties in relation to the principle of informed consent. Insofar as such research does not involve the research subjects' abdication of control over the research process, it may be less of an issue than in biomedical research. On the other hand, the very unpredictability of qualitative research design may mean that it is impossible to achieve fully informed consent at the beginning of a study.

These complexities suggest that any attempt to extrapolate practices designed to safeguard ethical research in biomedicine to qualitative studies mechanically is likely to be unhelpful. The relative standardisation of biomedical research makes the development of standardised practices for protecting research participants feasible. In qualitative research the situation is markedly different. We would argue that the same ethical principles (autonomy, protection of the research participant from harm, justice) apply in both kinds of research. However, in the context of qualitative research, these cannot be guaranteed simply by insisting that all participants sign a consent form or by a simple weighing of anticipated risks and benefits. Rather, they require, first, that the researcher should take responsibility, in advance of the study, for reflecting upon the possible ethical implications of the proposed work and considering whether, in the light of these, the research may proceed at all. Second, the researcher should consider the ways in which all residual risks to research participants may be reduced to a minimum. Such measures may include technical procedures for the anonymisation of data, similar to those used in biomedical research. They may also include consideration of how any risk of psychological or emotional harm to participants may be reduced and how any potentially negative consequences of publication of the study may be contained. Finally, the researcher must give attention to how the requirements of informed consent may be met, not just at the start of the study, but also as the research design emerges in the course of the study.

Implications for commissioning and practice

- The generalisability of qualitative research in HTA is extremely important, as commissioners and users are normally less interested in particular settings for their own sake than in the extent to which findings from one setting can be extrapolated to others. If the findings of qualitative research are to be relevant to the concerns of practitioners and policy makers then issues of sample selection, representativeness and generalisability must be addressed in research proposals and reports.
- There is no reason, in principle, why probabilistic sampling methods should not be used in qualitative research. However, in practice, such methods are rarely efficient or cost-effective. In qualitative research the use of non-probability sampling methods is best treated as a trade-off between depth and breadth.
- In selection decisions in qualitative research in HTA, pragmatic considerations should be integrated with a commitment to drawing samples in a systematic and principled way. Opportunistic sampling should be avoided in all but the most exploratory research and in some exceptional studies of stigmatised behaviour or conditions (e.g. HIV/AIDS).
- Where representativeness or typicality is the central concern, qualitative researchers should make use of purposive sampling strategies to foster the empirical generalisability of their findings. Such strategies include identifying settings or groups known to be typical in some way of the aggregate of interest. In some cases it will be appropriate to select a setting or group that is known to be atypical in order to test, delimit or extend the generalisability of findings.
- Where the concern is to build or develop theory, sampling should be directed towards making possible generalisations to theoretical propositions. Cases should be selected to enable researchers to test some element of the developing theory. Such theoretical sampling makes use of existing theory to make predictions and then seeks cases which allow the researcher to test the robustness of such predictions under different conditions. As such it parallels the use of critical experiments in the laboratory. In HTA, theoretical sampling can enable researchers to take findings which have been established in one context and test their applicability in another.
- The emergent nature of qualitative research design means that sampling decisions need to be made recurrently throughout the study, as well as at the outset. Such recurrent decisions will include who, what, when and where to study within the setting(s) or groups initially selected. Such decisions should again be systematic and principled. The study of negative, critical, discrepant and deviant cases is particularly useful in establishing the limits of the assertions which researchers seek to make on the basis of their observations.

- Participant observation typically involves combining periods of first-hand observation with formal and informal interviews and documentary analysis.
- Participant observation can be used in HTA to study the impact of technologies upon the routine, everyday functioning of the setting in which they are to be implemented.
- Participant observation can be used as a precursor to intervention studies to enable the researchers to specify commensurable categories of practice to permit the standardisation of interventions required by experimental logic. Such preliminary research is particularly important where, as is often the case, the intervention is not a drug-like entity which can be standardised easily. Such observational research will help to avoid the premature operational definition of variables in interventions which undermine the validity and applicability of findings.
- Participant observation can be used alongside intervention studies to provide a detailed description of the context in which, and the process through which, an intervention was implemented. Given that it is rarely possible to achieve the degree of control possible in the laboratory in studying the implementation of health technologies, the specification of such contexts and processes is necessary for the generalisability of findings.
- Participant observation can also be used to explain unanticipated or inconclusive findings from quantitative studies. In particular, where the introduction of health technologies in applied settings fail to produce the benefits hypothesised by basic the basic clinical sciences, such studies can uncover contextual factors which are responsible.
- Participant observation studies are also appropriate for the study of the unintended consequences of the introduction of new health technologies.
- Participant observation studies may be used to review health technologies currently in use. By studying the practices of health professionals or the interactional encounters between professionals and their patients/clients, observational studies in healthcare settings have the potential for uncovering the processes through which professional inputs are transformed into patient/client outcomes, identifying opportunities for modifying current practice to improve outcomes.
- In participant observation, the focus of the research is expected to narrow and data collection to become more targeted, as the study progresses. This approach, which is sometimes known as progressive focussing, allows the researcher to avoid the premature definition of variables.
- While more demanding, expensive and time-consuming than interviews, participant observation has substantial advantages in assessing the impact of health technologies. In particular, it frees us from the necessity of taking people's interview responses as reliable indicators of their likely behaviour. It avoids the problems of recall associated with interviews and allows the researcher to focus upon the mundane, everyday aspects of the settings in which health technologies are being practised or introduced.
- Careful consideration should be given in both research proposals and reports to the role played by the observer in observational research. The practicality, costs and benefits of taking on a participant role within the setting should be carefully weighed up. In particular, the researcher should demonstrate an awareness of the importance of marginality within the setting observed and the risks of being seduced by the perspective of one or other group (planners, professionals or patients) within the setting. The difficulty of maintaining such marginality should not be under-estimated and mechanisms, such as ensuring regular time away from the setting and peer-debriefing, should be in place.
- Qualitative interview techniques are used, particularly in exploratory research, to study the range and complexity of ideas and definitions employed by individuals and groups involved in the implementation of health technologies. Qualitative interviews may be used to follow up interesting ideas and to open up new dimensions, which had not been anticipated in advance of the interviews.
- Commissioners and researchers should, however, be wary of claims that qualitative interviewing gives access to participants' real understandings or feelings. Both qualitative and quantitative interviewing share the same fundamental problem that they rely upon interviewee's reports and such reports are necessarily the product of the situation in which they are offered. This undermines the use of interviews as literal descriptions of what planners, health professionals or patients think or intend.
- Interviews can, nevertheless, be analysed to uncover what those involved in administering or undergoing health technologies take to be self-evident. Once again, such information may be used to understand both factors which obstruct the implementation of such technologies and their unintended consequences.
- While observational studies have significant

advantages over studies that rely on interviews alone, not all research topics relevant to HTA are open to such observation. In particular, where health technologies are self-administered by patients, in geographically diverse locations, away from the semi-public settings of clinics, observational research may not be practical. In such cases, if the research is to be carried out at all it may be necessary to rely upon interviews. However, interview material should always be analysed in relation to the circumstances of its production.

- Written records proliferate in the settings in which health technologies are implemented and the analysis of such documents has an important contribution to make to our understanding of the processes and consequences associated with such technologies.
- In addition, documents such as diaries which are elicited specifically for research purposes, may provide important data on the implementation of health technologies.
- With increasing use of computerised technologies for keeping records in healthcare settings, detailed analysis of the functions of traditional record keeping in such settings can provide significant data about the shortcomings of current technologies, and the negative consequences of such innovations, and can inform the design of such innovations in ways that are likely to be acceptable to and implemented by practitioners.
- The use of documents as direct reproductions of reality raises the same problems as the use of interview material in this way. Such documents, whether naturally occurring or elicited specifically for research, are the product of the context in which they were created. As such, they must always be analysed in relation to the circumstances of their production. They may tell us more about the circumstances in which they were produced than about the events on which they purport to report.
- The techniques of CA can provide detailed data on the impact of new technologies upon healthcare settings and the interactions between health professionals and patients.
- HTA commissioners should look for evidence that applicants intend to use systematic methods for coding and handling their qualitative data and that the methods proposed for analysing such data are appropriate to the research objectives.
- In particular, such methods should include the systematic search for falsifying evidence, including, where appropriate, the identification and investigation of negative or deviant cases.
- Good qualitative research in HTA, as in other fields, will take account of existing research, in analysing the data, without imposing the findings of such earlier work on the analysis.
- Commissioners should be wary of researchers who claim to be using grounded theory or analytic induction without spelling out clearly what this will involve in relation to their specific project.
- Computerised analysis packages for qualitative data offer an efficient way of handling qualitative data sets and may improve the rigour of the analysis by facilitating searches for falsifying evidence. However, the analysis of qualitative data is not an algorithmic process and good qualitative analysis depends upon the theoretical sensitivity of the analyst. Thus, such packages must not be expected to carry out analyses. Rather they should be used as means of facilitating the analysis process.
- While the same ethical principles (autonomy, nonmaleficence, beneficence and justice) apply equally to both quantitative and qualitative research in HTA, the application of these principles should take into account the nature of the research proposed.
- The mechanical application of ethical codes, developed in the context of biomedicine, may be unduly constraining in qualitative research in HTA and may lead to the ritualistic observation of rules, at the expense of genuine ethical practice. Ethical practice in qualitative research cannot be guaranteed by insisting that all participants sign a consent form or by a straightforward calculus of costs and benefits.
- There are some ethical risks which are specific to qualitative research and the concentration upon the ethical risks associated with other forms of research may distract attention from these.
- The risks of serious physical harm arising from qualitative research are rarely of the same order as those which arise in some forms of biomedical research. However, such research may cause psychological harm, damage to self esteem and to interpersonal relationships. Qualitative researchers should seek to minimise such risks by making use of technical procedures for the anonymisation of data and seek to avoid any negative consequences of participation arising from the publication of results.
- All HTA research, both qualitative and quantitative, should only be carried out on consenting individuals. However, a signed consent form, typical of biomedical and psychological practice, does not always provide adequate or appropriate protection of the rights of participants in qualitative research. The qualitative researcher must ensure that the requirements of informed

consent are met, not just at the start of the study, but also as the research design emerges.

- Where experimental research involves the application of technologies to individuals, it is relatively clear who is at risk and, therefore, from whom consent should be sought. Qualitative research, on the other hand, will often involve assessing the impact of technologies upon groups rather than upon individuals and upon settings rather than upon subjects. In such cases, the researcher is obliged to consider from whom consent should be sought and to address the serious practical difficulties that arise when the membership of a setting or group is fluid or ill-defined.
- Covert research will rarely, if ever, be justified in HTA. Such covert research is likely to involve a betrayal of trust and a gross invasion of personal privacy.
- Confidentiality raises particular difficulties in relation to qualitative research where data may contain extraneous detail which might help to identify respondents.

Chapter 5

Criteria for assessing qualitative research

Commissioners of research have a particular interest in identifying criteria by which qualitative research may be assessed. The identification of such criteria is vital for the evaluation of both proposals to carry out research and of the findings of completed research. The methodological literature on qualitative research is littered with calls for more clarity on the criteria that are appropriate for qualitative research (e.g. Athens, 1984; Phillips, 1987), but the debate about whether criteria should be applied at all and, if so, which criteria, continues to be waged. This debate is complicated by issues discussed earlier in this review, including the lack of consensus about what qualitative research is, and the variety of approaches, each with its own rules, aims and logic, which are included in the term 'qualitative research' (Hammersley, 1990). The position was summed up by Sandelowski: "The debate surrounding the methodological rigour of qualitative research is confounded by its diversity and lack of consensus about the rules to which it ought to conform and whether it is comparable to quantitative research" (Sandelowski, 1986:29).

LeCompte and Preissle (1993c) suggested that this debate about criteria is of fairly recent origin. Until the 1970s, validity in qualitative research was discussed in relatively conventional terms. In the 1970s and 1980s some qualitative methodologists began to treat validity and reliability metaphorically rather than literally, focussing upon such issues as confidence, authenticity, cogency and soundness. By the 1980s, writers such as Lincoln and Guba (1985) had started to argue for the replacement of validity and reliability with criteria which, in their view, were more suited to qualitative research.[1] LeCompte and Preissle (1993c) suggested that, in the 1990s, this debate has become characterised by what they described as an **oppositional stance**, with some arguing that no common standards of evaluation are possible (Smith, 1984), others that criteria of goodness are paradigm specific (Lincoln and Guba, 1985; Guba and Lincoln, 1989), others that the validity of studies depends upon their success in revealing ideological distortions (Mies, 1983) and yet others adopting the more traditional criteria of validity and reliability (LeCompte and Goetz, 1982; Kirk and Miller, 1986; LeCompte and Preissle, 1993c). In this section, we shall lay out the range of positions adopted by qualitative researchers in relation to the criteria that are deemed appropriate for judging their work.

This debate can be seen as addressing, and sometimes conflating, three questions.

- Is it appropriate, or indeed possible, to identify criteria for evaluating qualitative research?
- If it is appropriate, what criteria should be adopted?
- Given the criteria adopted, how are these to be assessed in relation to any particular piece of research?

There are differences among qualitative researchers in relation to what they see as the aim of research as an activity, and hence what they identify and strive towards as good practice in qualitative research. Not all qualitative researchers accept the canons of rigour drawn from conventional research.

5.1 Is it appropriate and/or possible to identify criteria for evaluating qualitative research?

While most methodologists who write about qualitative research advocate the application of some standards by which good qualitative research may be distinguished from less good work, there are those who argue that the search for **any** such criteria is fundamentally misconceived. This position has been most clearly laid out by Smith (1984) and is rooted in the anti-realism[2] of the two-paradigms[3] position. From Smith's perspective, qualitative research represents a distinctive paradigm (Smith, 1985; Smith and Heshusius, 1986), and the defining characteristic of that paradigm is its commitment to relativism or anti-realism. This, Smith argued, involves accepting that reality is mind-dependent and the denial of any possibility

[1] See section 5.2; [2] See section 3.2.1.1 for a discussion of realism and anti-realism; [3] The two-paradigms perspective is discussed in section 3.1.2.

of a context-free reality which might be used as a criterion against which research findings might be judged. Smith argued that as people "may construct their realities in their own different ways at different times and places" (Smith, 1984:382), it is a nonsense to identify some versions of reality trustworthy and others not. Smith held that those, such as Lincoln and Guba (see below) who hold to relativism and yet attempt to develop criteria for assessing qualitative research are attempting the impossible:

> The assumptions of multiple realities and reality as mind-involved seriously undermine the notion of applying foundational criteria to distinguish trustworthy results from untrustworthy ones. The assumptions and foundational criteria are, in a word, incompatible ... different claims about reality result not from incorrect procedures but may simply be a case of one investigator's reality versus another's. (Smith, 1984: 383.)

Smith's argument is logically coherent. His conclusions follow from the premises he adopts. If we accept these conclusions, we are, however, forced to raise serious doubts about whether qualitative research has any contribution to make to HTA. If no standards can be applied to qualitative research, or to its findings, on what grounds could commissioners decide that it is appropriate to fund one piece of qualitative research rather than another? Indeed what justification is there for funding any research, whether qualitative or quantitative? Smith's response, that qualitative research is "an attempt to enlarge the conversation and keep it going" (Smith, 1984:390), is unlikely to satisfy commissioners or their paymasters. As Atheide and Johnson (1994) have argued: "As long as we strive to base our claims and interpretations of social life on data of any kind, we must have a logic for assessing and communicating the interactive process through which the investigator acquired the research experience and information" (485).

The force of Smith's argument depends crucially on his assumption that qualitative research is based on an anti-realist ontology and a relativist epistemology. If one does not accept this premise, then the argument that no criteria are possible falls. As discussed above, many qualitative researchers have rejected the association between qualitative research and relativism (Silverman, 1985; Phillips, 1990; Hammersley, 1992d; Hammersley, 1992e; Altheide and Johnson, 1994) and have proposed some form of, often modified, realism as an alternative basis for qualitative research.[4] Once the link between relativism and qualitative research is broken, one of the central premises of Smith's argument falls, and it becomes possible to consider the possibility of criteria for qualitative research.

5.2 What criteria are appropriate for qualitative research?[5]

Among the majority of qualitative researchers who accept the importance of identifying appropriate criteria for qualitative research, a number of distinctive positions can be identified. These positions have been classified in a number of ways (Goodwin and Goodwin, 1984a; Hammersley, 1990; Eisenhart and Howe, 1992; Hammersley, 1992d; Denzin and Lincoln, 1994a). For example, Hammersley (1990; 1992e) identified two broad positions on criteria, other than the total rejection discussed above. On the one hand, there are those who have argued that the same criteria (generally some version of validity and reliability) should be applied to qualitative and quantitative research, and that there is no distinctive philosophy underlying qualitative research (LeCompte and Goetz, 1982; Kirk and Miller, 1986; LeCompte and Preissle, 1993c).[6] On the other hand, there are those who have argued that qualitative research is a distinctive paradigm, which represents a different form of science and which, therefore, requires the application of different criteria (Morgan, 1983; Lincoln and Guba, 1985; Guba and Lincoln, 1989).[7] This second position, as Hammersley accepted, is not internally homogenous. There are, for example, those who have argued that research findings are to be judged in terms of their capacity for producing valued change (Owens, 1982; Mies, 1983), in terms of their isomorphism with participants' perspectives (Lincoln and Guba, 1985; Guba and Lincoln, 1989) or in terms of their aesthetic value (Tesch, 1991).

Goodwin and Goodwin (1984a) have proposed an alternative, fourfold classification of the positions adopted by qualitative researchers in relation to the traditional criteria of validity and reliability for assessing their work. First, there are those, such as Smith (1984), discussed above,[8] who see validity and reliability as irrelevant. Second, there are those who argue that validity is important for qualitative

[4] See section 3.2.1.1; [5] See section 2.3 for the historical background to this debate; [6] See section 3.2.1.1; [7] See section 3.1.2; [8] See section 5.1.

research but that reliability is irrelevant (Reichardt and Cook, 1979; Harding, 1987). Third, there are those who hold that reliability and validity are both important but difficult to establish in qualitative research (Bogdan and Biklen, 1982). Finally, there are those who argue that both reliability and validity are important and can be studied directly (Denzin, 1970; LeCompte and Goetz, 1982).

Denzin and Lincoln (1994a) presented a further alternative classification of positions on criteria for evaluating qualitative research.

- Those who want to apply the same criteria to qualitative and quantitative inquiry.
- Those who argue that a set of criteria specific to qualitative research needs to be developed. This is usually couched in a commitment to the idea that qualitative research represents a distinctive methodological paradigm.[9] This position extends to include constructivists who propose trustworthiness and authenticity as criteria and to critical theorists who propose **action**, **praxis** and the **historical situatedness of findings**.[10] This position also incorporates those who advocate such criteria as, "emotionality, caring, subjective understanding, dialogic texts and the formation of long term trusting relationships with those studied" (480).
- Those who adopt a postmodern position such as that of Smith discussed above, which "doubts all criteria and privileges none" (480).
- Those who adopt a poststructuralist position, which contends that "an entirely new set of criteria, divorced from the positivist and postpositivist traditions, needs to be constructed" (480).

In presenting and assessing the arguments associated with the various positions outlined above, we have chosen to simplify this complex picture, by comparing the arguments of those who advocate the application of the same or broadly similar criteria to all research, regardless of whether it is qualitative or quantitative, with those who argue that the distinctive features of qualitative research call for an alternative set of criteria for its evaluation. The latter position can be seen as closely associated with those who argue that qualitative and quantitative research should be seen as two incommensurable paradigms (Guba, 1981; Morgan, 1983; Lincoln and Guba, 1985; Marshall, 1985; Sandelowski, 1986; Guba and Lincoln, 1989).[11] The former is associated with postpositivism (Phillips, 1990; Denzin and Lincoln, 1994b; Guba and Lincoln, 1994; Seidel and Kelle, 1995).[12] We shall deal first with those who have argued for distinctive criteria to be applied to qualitative research.

5.2.1 Advocates of alternative criteria for qualitative research

Marshall (1985) has criticised the wholesale application of validity and reliability to qualitative research on the pragmatic grounds that imposing such criteria runs the risk of constraining the work of discovery and hypothesis generation, which she sees as one of the strengths of qualitative method.[13] Unlike many of those who resist the extension of traditional quantitative criteria to qualitative work, Marshall does not appear to be concerned here about ontological or epistemological issues. She identified two separate strands within qualitative research. For the first, testing hypotheses in context, the application of traditional criteria, was seen as perfectly appropriate (see section 5.2.2.). It is in relation to the second strand, that of exploration, discovery and hypothesis generation, that Marshall suggested that such criteria are overly restricting and run the risk of constraining qualitative research to valid but insignificant parameters.

While Marshall sought to distance herself from what she described as the positivist criteria of validity and reliability, she nevertheless wished to see standards applied to qualitative research. When one examines the means of achieving the standards which she proposed, it becomes clear that it is not so much the criteria which she objected to, but the means that are traditionally used to apply such standards to quantitative research. In particular, it is the emphasis on control and systematisation which she presented as constraining the possibility of discovery within qualitative research and the imposition of *a priori* theoretical frameworks,[14] which she saw as compromising the hypothesis generation potential of qualitative research. This confusion highlights the importance of distinguishing between the criteria that are applied and the means which are used to assess such criteria (Hammersley, 1990).[15]

Lincoln and Guba (1985) represent a more radical critique of the application of the criteria derived

[9] See section 3.1.2; [10] See section 3.1.3; [11] See section 3.1.2; [12] See section 3.1.1; [13] See sections 3.2.1.2 and 3.2.2.5; [14] See section 3.2.1.2; [15] In this section of the review we are focussing upon identifying the criteria that should be used to judge the quality of qualitative research. In sections 5.3 and 5.4, we shall discuss the means by which research may be assessed in relation to these criteria.

from the so-called **quantitative paradigm** to qualitative research, than that proposed by Marshall. They argued that it is fundamentally inappropriate to apply the criteria derived from one paradigm to the products of research grounded in an alternative paradigm. They presented any universal criteria for evaluating research as inevitably flawed and bound to favour research strategies that are congruent with the assumptions of the paradigm from which they are derived. They took four criteria, which they believed to be characteristic of the positivist, quantitative paradigm, and offer alternatives which they see as more appropriate to the qualitative paradigm.[16]

5.2.1.1 Validity

Lincoln and Guba (1985) argued that conventional concern with validity as a means of evaluating research is based on the assumptions of naive realism. Like Smith (1984),[17] Lincoln and Guba rejected the idea that there is a single reality in favour of an assumption that reality is always constructed (or mind-dependent, Smith, 1985). Given this assumption of multiple, constructed realities, they argued that there can be no ultimate benchmark for judging the truth value of any claim. They rejected all claims to validity that are based on isomorphism between research findings and reality. They argued: "If realities are instead assumed to exist only in mentally constructed form, what sense could it make to look for isomorphisms?" (Guba and Lincoln, 1989:236).

Lincoln and Guba suggested that qualitative researchers should replace the positivist criterion of validity, with that of 'credibility'. Within their scheme, credibility has a somewhat restricted meaning. It refers to the endorsement of researchers' conclusions by the subjects of the research. Thus, a researcher's conclusions are to be deemed credible if the constructors of the particular perspectives being reported confirm that they represent their particular version of reality. The findings of qualitative research are to be assessed against the various sources from which these findings have been drawn. The problems associated with such 'member validation' approaches to assessing the quality of qualitative research are discussed in section 5.3.1.

5.2.1.2 Generalisability or external validity

Lincoln and Guba (1985) also rejected traditional concerns with generalisability or external validity as inappropriate within the qualitative research paradigm. This once again reflects their rejection of realist assumptions: "External validity, a concept that embodies the very essence of generalizability, likewise can have little meaning if the 'realities' to which one might wish to generalize exist in different forms in different minds, depending on different encountered circumstances and history, based on different experiences, interpreted within different value systems" (236).

Whereas, in quantitative research, the demands of external validity require research to be carried out in such a way that time and situational variables are irrelevant to findings and truth statements are context free, within the qualitative paradigm there is, according to Lincoln and Guba, acceptance that such generalisations are impossible because phenomena are intimately tied to the times and contexts in which they are found. Once again, they associated this criterion of external validity with the realist assumption that nature is rule governed in all circumstances and that the findings of research in one setting should be directly applicable to other parallel settings. They propose **transferability** as an alternative criterion that is more appropriate to qualitative research. Although direct comparability between settings is impossible, some similarities do exist between different settings and it is possible to develop working hypotheses which have some potential for transfer between different settings. Being able to evaluate the possibility of such transfer depends crucially upon the researcher providing 'thick description' (Geertz, 1973) of the original setting. Thick description as a means of assessing the transferability of qualitative research will be discussed in section 5.4.

5.2.1.3 Reliability

The traditional emphasis on reliability was rejected by Lincoln and Guba, on the grounds that it reflects realist[18] assumptions, which are held to be inappropriate to the qualitative paradigm (see also Merriam, 1988). Lincoln and Guba (1985) and Guba and Lincoln (1989) argued that traditional approaches to assessing reliability assume an unchanging reality which can be used as a benchmark. By contrast, they suggested, the qualitative researcher assumes that the entity being studied **will** change. The assumption of multiple realities is taken to mean that observations change not only because of error but also

[16] As Hammersley (1990) has pointed out there is not complete agreement among quantitative researchers about the criteria which should be used to judge their research. However, two sets of concepts (internal and external validity; validity and reliability) are commonly drawn upon. These are discussed more fully below; [17] See section 5.2; [18] See section 3.2.1.1. for a discussion of realism.

because of changes in the researcher's insights and sensitivities. It is, therefore, seen as important that researchers accept that at least some portion of the instability which they identify is recognised as real rather than a methodological artefact. They proposed the alternative criterion of consistency or dependability. They defined this criterion as **trackable variance**, which is attributed to sources. For a study to be regarded as dependable, it will be necessary for researchers to have taken into account the inherent instability of the phenomenon they are studying and the way in which the research design may have induced change in that phenomenon.

Merriam (1988) argued: "Because what is studied is assumed to be in flux, multi-faceted, and highly contextual, because information gathered is a function of who gives it and how skilled the researcher is at getting it, and because the emergent design of a qualitative case study precludes *a priori* controls, achieving reliability in the conventional sense is not only fanciful but impossible" (171).

5.2.1.4 Neutrality

Lincoln and Guba (1985) and Guba and Lincoln (1989) argued that the traditional criteria of neutrality or objectivity in research, which rest on assumptions about the possibility of value freedom, should be replaced by confirmability. Judging the findings of research confirmable depends upon the provision of an **audit trail**, which would allow other researchers to examine the process by which researchers have arrived at their conclusions. This shifts the emphasis away from the concept of investigator objectivity to the concept of data and interpretational confirmability.

Lincoln and Guba's position is taken up in a number of subsequent publications, though later writers often apply different labels to similar ideas. For example, in the context of an article, which argued for the complementarity of qualitative and quantitative approaches in nursing research, Beck (1993) drew upon Lincoln and Guba to suggest that appropriate criteria for qualitative studies include credibility in place of validity, fittingness in place of generalisability, and auditability in place of reliability. As for Lincoln and Guba, for Beck a study will be credible if the researcher's description is recognised by informants and readers as a faithful rendition of their own experiences.[19] Fittingness is parallel to Lincoln and Guba's criterion of transferability and concerns the extent to which the findings are deemed to fit settings beyond the one that has been studied. Auditability concerns the extent to which another investigator can follow the **decision trail** which led the original researchers to their conclusions.

Similarly, again from a nursing perspective, Sandelowski (1986) criticised the tendency to evaluate qualitative research against criteria that are more appropriate to quantitative research, and advocated a position which emphasises the relative nature of truth. She adopted Lincoln and Guba's criterion of credibility in preference to the conventional criterion of validity on the grounds that the former deals with truth as subject-orientated rather than researcher-defined. She argued for fittingness rather than representativeness as an appropriate criterion for qualitative research, in recognition of the ultimate uniqueness of all settings. She preferred auditability to reliability, on the grounds that reliability depends upon a series of unwarrantable assumptions, including that there is an observable regularity about human experiences and that if more than one person observes a phenomenon it exists in the same way for each of them. Finally, she rejected the quantitative criterion of objectivity as inappropriate for qualitative research, which emphasises the importance of reducing the distance between the researcher and the researched and accepts that there is no possibility of studying reality without changing it in the process. Rather than seeking to overcome subjectivity, qualitative researchers are encouraged to exploit it.

Leininger (1992) also linked her rejection of conventional criteria to her insistence that qualitative and quantitative research are grounded in two incommensurable paradigms[20] and that attempting to mix methods from different paradigms "violates the purposes and integrity of the paradigms". She put her case forcefully: "To make quantitative criteria **fit** qualitative purposes is analogous to using a tape measure to understand the meaning of a concern to individuals or groups" (402).

Somewhat confusingly, Secker and co-workers (1995) described this kind of anti-realist approach as sociological, without apparently realising that sociological research spans the entire spectrum from naive realism to radical relativism.[21] In the context of health education, these authors argued that both realist and anti-realist approaches should

[19] See section 5.3.1 for a critique of respondent validation; [20] See section 3.1.2. for a discussion of the two-paradigms argument; [21] See section 3.2.1.1. for a discussion of realism and relativism.

be employed on a 'horses for courses' basis, and that the products should be assessed on the basis of criteria which are appropriate for the particular approach. When adopting a **sociological** approach, researchers should not be concerned with amassing objective facts about the world. Rather the research product should be recognised as the researcher's interpretation, rather than a reflection of reality. In common, with many of those writing from an anti-realist perspective, there is a contradiction between their assertion that research findings can be no more than a researcher's interpretation and their subsequent attempt to identify means of assessing the 'goodness' of such interpretations. On what grounds are we to assess the goodness of an interpretation if it is not in relation to some criteria which are external to that interpretation?

From the perspective of psychology, Henwood and Pidgeon (1993) again argued for the distinctiveness of the qualitative paradigm. While they emphasised the importance of not exaggerating the epistemological differences between qualitative and quantitative research, they nevertheless rejected the appropriateness of traditional criteria for assessing qualitative research. In the light of their insistence that qualitative research is grounded in a constructivist epistemology, they rejected traditional criteria (such as reliability, validity, internal consistency and generality) as dependent upon a scientific norm of objectivity, directed towards the goal of eliminating bias, where bias is understood as deviation from objective truth. Henwood and Pidgeon argued that the personal is always present in research and the evaluation of research involves more than establishing that the researcher has succeeded in eradicating bias.

The influence of Lincoln and Guba's early writings (Guba, 1981; Lincoln and Guba, 1985) can be detected in all of those authors discussed so far. However, as LeCompte and Preissle (1993c) observed, the position which Lincoln and Guba adopted in these early writings can be seen as a staging post to the more radical position, which they later adopted (Guba and Lincoln, 1989). Indeed, their early work has been subject to the criticism that their attempt to adopt the relativist position, that there are multiple and conflicting truths which can nevertheless all be true, while at the same time seeking to develop and apply criteria of warrantability is untenable (Smith, 1984; Phillips, 1987). Smith (1984), discussed above,[22] summarised this critique:

> Guba's work therefore provides an excellent example in tangible terms of what happens when antifoundational assumptions are combined with the attempt to pose foundational criteria. The inevitable result is inconsistency of argument. (384)

In their later work, Guba and Lincoln did not abandon their earlier attempts to identify **parallel** criteria for assessing research within the constructivist paradigm (Guba and Lincoln, 1989). However, they acknowledged the force of critics such as Smith (1984), discussed above, and suggested an additional set of criteria which they see as more in keeping with their relativist assumptions:

> A relativist ontology asserts that there exist multiple, socially constructed realities, ungoverned by any natural laws, causal or otherwise. 'Truth' is defined as the best informed ... and most sophisticated ... construction on which there is consensus (although there may be several constructions extant which simultaneously meet that criterion). A monist, subjectivist epistemology asserts that the inquirer and the inquired into are interlocked in such a way that the findings of an investigation are the **literal creation** of the inquiry process. (Guba and Lincoln, 1989:84 original emphasis.)

In the light of this more fully developed relativist position, they proposed five alternative criteria for assessing the goodness of evaluation research, which they believed to be compatible with relativism. They defined these as criteria of **authenticity**, which they discussed particularly in the context of evaluations. They are summarised below.

- **Fairness:** researchers must be able to demonstrate that they have represented the range of different realities in a balanced way.
- **Ontological authenticity:** researchers must be able to demonstrate that, in the course of the evaluation, members have developed a more sophisticated understanding of some phenomenon, than they possessed at the outset.
- **Educative authenticity:** researchers must be able to demonstrate that in the course of the evaluation, members have developed a greater understanding of and appreciation for the understandings of other members or groups.
- **Catalytic authenticity:** researchers must be able to demonstrate that the evaluation process has stimulated action.
- **Tactical authenticity**: researchers must be able to demonstrate that members have been empowered to act.

[22] See section 5.1.

Similarly, Eisner (1983) offered three criteria for evaluating qualitative research, which are in line with a relativist position.

- **Structural corroboration:** by this Eisner means that research accounts, drawing upon different parts of the data, should demonstrate coherence.
- **Referential adequacy:** readers should be presented with data which enable them to 'see what the researcher is talking about'.
- **Multiplicative replication:** the members of the community should believe the findings.

As will be apparent from this discussion, even among those who believe that distinctive criteria should be developed for qualitative research, there is substantial variability in the criteria that are proposed. Hammersley (1990) attempted to draw these together into a composite list of criteria, which have been proposed by a number of authors (Frake, 1962; Wolcott, 1975; Athens, 1984; Miles and Huberman, 1984; Lincoln and Guba, 1985; Lather, 1986; Hammersley, 1987).

1. The degree to which substantive and formal theory are produced and the degree of the development of the theory.
2. The novelty of the claims made.
3. The consistency of the claims with empirical observations.
4. The credibility of the account to readers and/or to those studied.
5. The extent to which the cultural description produced provides a basis for competent performance in the culture studied.
6. The extent to which the findings are transferable to others settings.
7. The reflexivity of the account – the degree to which the effects of research strategies on the findings are assessed and/or the amount of information about the research process that is provided to readers (56).

Hammersley identified a number of problems with these criteria. He rejected the assumption, upon which the first criterion depends, that all research is, or should be, concerned with producing theory. He also argued that a number of these criteria (3, 4, 5, 7) are concerned with the **means** of assessment rather than the standards against which assessment should be made (see sections 5.3 and 5.4) Finally, he argued that the claim that these criteria are specific to qualitative research is unfounded, since they raise issues of importance however the data are collected.

Alternative criteria have also been proposed by those working from the perspective of critical theory (Eisner, 1983; Marshall and Rossman, 1989; Roman and Apple, 1990).[23] For example, Roman and Apple (1990), from a feminist, materialist perspective have rejected attempts by both experimental and qualitative researchers to achieve validity through limiting the researcher's involvement. Such approaches are seen as fundamentally flawed as they do not question the power relations of the wider society and the ways in which these are perpetuated in research. They argued that studies should be democratically designed and the results should be democratically produced. Validity is defined in terms of the extent to which such democratisation is achieved. On this basis, they proposed four criteria for judging the validity of qualitative research:

- the extent to which the study resonates with the 'lived experience' of those being researched
- the extent to which the study enables members of the group being studied to comprehend and transform their experiences of subordination
- the extent to which the study reduces the divide between the researcher's intellectual work and group members' ordinary ways of describing and understanding their experiences.
- the extent to which the research allows the researcher's prior theoretical and political commitments to be informed and transformed by understandings derived from the group's experiences (63–4).

Similarly, Mies (1983) has suggested that the validity of research is to be judged in terms of its contribution to the emancipation: "... the 'truth' of a theory is not dependent on the application of certain methodological principles and rules, but on its potential to orient the processes of praxis towards progressive emancipation and humanization" (124).

Greene (1996) argued that research should be evaluated in terms its social, political and value consequences. She rejected the notion that the goodness of research should be seen primarily in terms of ontology, epistemology or methodology. Appropriate criteria should include politics, morals and philosophy as well as ethics.

[23] See section 3.1.3.

5.2.2 Advocates of the application of conventional criteria from quantitative research to qualitative research

Many of the reservations about conventional approaches to judging research and the alternative criteria proposed for qualitative research, which were reviewed in the previous section arise from a commitment to an anti-realist position.[24] As we have suggested, there are significant problems with this perspective, particularly for those who wish to use research to inform policy and practice. However, qualitative research need not necessarily be associated with such anti-realism. For example, Hammersley (1990; 1992d; 1992e) argued that naive realism is not the only alternative to anti-realism. As discussed in section 3.2.1.1., the central tenet of naive realism is the belief that there is a reality, which is independent of the research which can be known, if only we can rid ourselves of the cultural preconceptions which distort our observations of it. As Hammersley argued, this position fails to recognise that all perceptions and observations are assumption-laden. However, this does not mean that we are forced to adopt the alternative of naive relativism:

> There is a great danger of backing ourselves into a corner by deploying a dichotomy which obscures the wide range of epistemological positions available. We can maintain belief in the existence of phenomena independent of our claims about them, and in their knowability, without assuming that we can have unmediated contact with them and therefore that we can know with certainty whether our knowledge of them is valid or invalid. The most promising strategy, in my view, then, is to adopt a more subtle form of realism. (Hammersley, 1992c:50.)

Hammersley identified three key elements of such subtle realism.[25]

- The idea that truth applies to knowledge, about which we are certain beyond any possible doubt, is rejected. Knowledge can never be certain in this sense because every claim to validity necessarily depends upon assumptions, the validity of which we must presuppose (Hammersley, 1990; Steckler *et al,* 1992). Rather Hammersley argued that truth should be reinterpreted as "beliefs about whose validity we are reasonably confident" (Hammersley, 1992c:50). Judgements about the validity of claims can only be made on the basis of the compatibility of such claims with "assumptions about the world that we currently take to be beyond reasonable doubt" (51).
- The belief that there are phenomena, which are independent of our claims about them and which our claims may represent more or less accurately, is upheld. By independent, Hammersley meant that simply making a claim does not in itself change relevant aspects of reality so as to render claims either true or false.
- The aim of social research is seen as representing rather than reproducing reality. This element accepts the argument that phenomena can be represented from **multiple perspectives**. However, this is not the same as the anti-realist argument that there are **multiple realities** (Phillips, 1987). Representation is always from some point of view and this means that certain features of a phenomenon will be treated as relevant and others as irrelevant. "There can be multiple, non-contradictory and valid descriptions and explanations of the same phenomena" (Hammersley, 1992d:51).

Other writers have argued for a similar modified realist position, variously identified as **fallible realism** (Greene, 1996) or **analytic realism** (Altheide and Johnson, 1994), or **critical realism** (Campbell,1994). Kirk and Millar (1986) pointed to the gap between their own position and that of the anti-realists discussed above:

> There is a world of empirical reality out there. The way we perceive and understand that world is largely up to us, but the world does not tolerate all understandings of it equally (so that the individual who believes that he or she can halt a speeding train with his or her bare hands may be punished by the world for acting on that understanding).(11)

This escape from relativism and the **two incommensurable paradigms** approach to social research, is associated with the argument that the same criteria should be used to evaluate both qualitative and quantitative research. This argument takes two forms (Hammersley, 1990).

On the one hand, it is argued by some that the conventional criteria by which quantitative research are judged (usually conceptualised as some combination of internal validity, external validity, reliability and construct validity) can and should be applied to qualitative research (Cook and Campbell, 1979; LeCompte and Goetz, 1982; Goodwin and Goodwin, 1984a; Kirk and Miller, 1986; Merriam, 1988; Jensen, 1989; LeCompte and Preissle, 1993c), usually suitably modified to take into account the design-

[24] See section 3.2.1.1; [25] See section 3.2.1.1 for a discussion of subtle realism. Here we consider the implications of this position for the evaluation of qualitative research.

specific features of particular qualitative methods (Eisenhart and Howe, 1992). On the other hand some authors, notably Hammersley (1990; 1992e), have argued that while the same criteria should be applied to both qualitative and quantitative research, there are some problems with the conventional criteria of internal and external validity and reliability. We shall outline the argument that traditional criteria should be modified for qualitative research, before going on to deal with Hammersley's proposed modification.

5.2.2.1 Applying conventional criteria to qualitative research

LeCompte and Goetz (1982) and LeCompte and Preissle (1993c) are typical of those who argue for the application of conventional criteria to qualitative research. They proposed that all researchers, whether within the qualitative or quantitative traditions, must strive for validity and reliability. However, they believed that: "Some factors confounding the credibility of findings in experimental research are inapplicable to ethnographic research; others need to be defined in special ways" (32).

These conventional criteria were developed in relation to experimental research and, in applying them to qualitative research, it is necessary to modify (or translate) them. They identified three areas in which qualitative and experimental research differ in ways which have implications for the application of criteria of reliability and validity to qualitative research.

Formulation of problems. In experimental research the concern to examine the effect of specific treatments, means that the emphasis is upon holding constant or eliminating contextual or extraneous factors. In qualitative research, on the other hand, the emphasis is upon examining the interplay among variables within a natural context.[26] Here the emphasis is upon systematically identifying and examining all causal and consequential factors. In qualitative research the focus is often upon the interrelationship among factors which in experimental research one might wish to hold constant.

Nature of the goals of research. In experimental research the goal is to test causal propositions which have been developed externally to the research site. The goals of qualitative research are more complex and vary according to the stage of the research process. LeCompte and Goetz (1982) listed the various goals of qualitative research, which may vary between different studies and within any one study over time: "Ethnographers attempt to describe systematically the characteristics of variables and phenomena, to generate and refine conceptual categories, to discover and validate associations among phenomena, or to compare constructs and postulates generated from phenomena in one setting with comparable phenomena in another setting" (33).

Application of results. The findings of most experimental and survey research are intended to be generalised to from the research subjects or settings to some wider population. Such researchers justify their claims to generalisability on the basis of the statistical sampling methods used,[27] design controls and assumptions of equivalence. Qualitative researchers are rarely able to use these techniques[28] and, as a consequence, they aim for comparability (based on the careful delineation of the characteristics of those studied or the constructs generated) and/or translatability (explicit identification of research methods, analytic categories and characteristics of phenomena and groups).

These authors then went on to suggest ways in which each of the conventional criteria should be modified to take account of the special features of qualitative research as follows:

Reliability. LeCompte and Goetz (1982) suggested that reliability, as conventionally defined ('the extent to which a researcher using the same methods can obtain the same results as those of a prior study'), poses a Herculean problem for researchers concerned with naturalistic behaviour or unique phenomena (see also Duffy, 1985). Several features of qualitative research, in particular, contribute to the difficulties in replication. First, qualitative researchers' preference for carrying out research in natural settings makes exact replication impossible insofar as unique situations cannot be precisely reconstructed (see also Jensen, 1989) and people's behaviour cannot be subjected to the same degree of control as in research on other animals or inanimate objects (Hammersley, 1990). Second, since the focus of much qualitative research is upon process and change (Dingwall, 1992),[29] the assumption of stability which underlies the notion of reliability is clearly problematic (see also Atwood and Hines, 1986).

LeCompte and Goetz (1982) and LeCompte and Preissle, (1993c) argued that, in considering reliability, it is important to distinguish between internal

[26] See section 3.2.2.3; [27] See section 4.1; [28] See section 4.1; [29] See section 3.2.2.4.

reliability (i.e. the extent to which, given a set of previously generated concepts, new researchers would match these concepts with the data in the same way as the original researchers) and external reliability (the extent to which independent researchers would discover the same phenomena or generate similar concepts in the same or similar settings). In terms of this distinction, it is external reliability which poses particular problems for qualitative research, and LeCompte and Goetz (1982) concluded that, given the difficulties outlined above, qualitative research can only approach, not attain external reliability. In section 5.3, we summarise some of the means by which qualitative researchers may seek to enhance the reliability of their data.

Internal validity. LeCompte and Preissle (1993c) defined validity as the extent to which propositions generated, refined and tested match what occurs in human life. Qualitative researchers often claim that the very nature of the data collection and analysis methods which they use give them significant advantages over quantitative research in terms of validity (Denzin, 1970; Schatzman and Strauss, 1973; Atwood and Hinds, 1986; Kirk and Miller, 1986; Duffy, 1987; Merriam, 1988; Guba and Lincoln, 1989; Mechanic, 1989). Critics of quantitative research sometimes argue that quantitative research typically pursues reliability at the expense of validity. Atwood and Hinds (1986) suggested that techniques such as the constant comparative method,[30] which are commonly employed in the analysis of qualitative data, bring a strong validity assurance because data are used to guide the conceptual development of labels and definitions for categories and properties. This means that validity checking may be continuous throughout the analysis process. Marshall and Rossman (1989) argued that: "An in-depth description showing the complexities of variables and interactions will be so embedded with data derived from the setting that it cannot help but be valid" (145).

Becker (1970), cited in Emerson (1981), identified two features of field work that promote high internal validity.

- The data is less artifactual than most other research data.
- As large numbers of observations are made over extended periods of time, findings are recurrently revised and revisited.

LeCompte and Goetz (1982) identified four features of qualitative research that enhance the validity of findings.

- Prolonged involvement in the field allows researchers to constantly refine their concepts and to ensure that researchers categories match participant realities (see also Duffy, 1985). Kirk and Miller (1986) argued that such prolonged engagement increases the validity of qualitative research by increasing the possibility that researchers' initial and possibly erroneous assumptions will be exposed to challenge. In addition, the preference, in qualitative research, for multiple methods is held to increase the validity of findings.
- Qualitative interviewers seek to avoid the risks of invalidity, which arise when researchers seek to impose their categories on informant thinking, rather than working with categories which are derived from the informants themselves.[31]
- The focus on conducting research in naturally occurring settings means that findings are likely to reflect the "reality of the life experiences of participants more accurately than do contrived settings" (43).[32]
- Qualitative analysis incorporates a process of researcher self-monitoring which exposes the research process and findings to continual questioning and reevaluation.

While LeCompte and Preissle (1993c) appear to have accepted the argument that qualitative research is inherently more likely to yield valid results than more quantitative methods, they nevertheless proposed that the findings of qualitative research should be assessed in relation to the same threats to validity which Campbell (1957) proposed for quasi-experiments. He listed particular threats to internal validity (history and maturation, observer effects, selection and regression, mortality and spurious correlations) and discussed the ways in which these relate to qualitative research. He argued that an assessment of the internal validity of a particular piece of qualitative research depends upon the extent to which the researcher has recognised all potential threats to validity and taken these into account in the design and analysis of the study.

External validity. LeCompte and Goetz (1982) argued that external validity is too often ignored by qualitative researchers. They attributed this to the tendency to focus on a single setting and the preference for suspending preconceived notions and existing knowledge of the field under study. As

[30] See section 4.7; [31] See section 3.2.1.2; [32] See section 3.2.1.3.

qualitative research rarely uses probability sampling techniques, statistical generalisation from one study to another is rarely possible.[33] However, LeCompte and Preissle (1993c), argued that qualitative researchers should nevertheless be concerned about the external validity of their findings. They suggested that, once again, qualitative researchers should concern themselves with the threats to external validity, identified by Campbell (1957) (selection effects, setting effects, history effects and construct effects).

5.2.3 The limitations of conventional criteria

While Hammersley (1990; 1992e) agreed with those authors, discussed above, that the same criteria should be applied to qualitative and quantitative research, he highlighted some of the limitations of conventional criteria. He acknowledged that these attempts to apply criteria drawn from the quantitative tradition to qualitative research help to identify important considerations, but he concluded that they fail to provide "a clear and coherent, or sufficient, conceptual basis" (Hammersley, 1990:55) for the assessment of research. The flaws which he identified apply to quantitative research as well as to qualitative work.

Hammersley (1990) outlined two sets of concepts in terms of which research conventionally is assessed and pointed to shortcomings in each. First, he examined Campbell and Stanley's distinction between internal and external validity (Hammersley, 1990). He argued that this distinction is based upon the untenable assumption that it is possible to separate the claim that a causal relationship has been found in one setting from the claim that it will be found in other cases of the same type. Hammersley's point was that the very concept of cause implies that the relationship observed will either always or at least probabilistically hold when the appropriate conditions are met. He concluded that the distinction between internal and external validity is therefore misleading. Turning to the second set of concepts, reliability and validity, Hammersley identified further problems. He pointed out that these terms are drawn from the domain of measurement theory and, as a result, cannot provide a sufficient set of assessment criteria, even for quantitative studies. In addition, he identified conceptual problems with these criteria, which arise from the diverse ways in which the terms are used. On the one hand, they can be used to refer to properties of measurement instruments, of observers or of particular measurements. On the other, they may be defined in terms of the relationship between findings and what is being measured or the relationship between findings produced by different instruments (Evans, 1983).

Hammersley (1990) argued that the confusion about what criteria are appropriate for assessing research arises partly from a failure to identify the function of research in advance of discussions about how it should be assessed. He proposed a function for research which is highly appropriate in the context of HTA: "To produce **knowledge** that is of public **relevance**" (56, original emphasis). From this definition of the function of research, he derived two criteria, against which such research may properly be assessed: validity and relevance. Hammersley defined validity as "truth: interpreted as the extent to which an account accurately represents the social phenomena to which it refers" (57). He accepted that absolutely certain knowledge is impossible:

> I believe that we can **never** be certain about the truth of anything, not even in the natural sciences or in our personal lives. On the other hand, there are many things about whose truth we are very confident and about which we have every right to be confident. (59 original emphasis)

His argument is that, while the validity of findings can never be established beyond doubt (see also Henwood and Pidgeon, 1993), it can and should be examined in terms of the likelihood of error. Various means for examining the likelihood of error of qualitative research (and hence assessing is validity) have been proposed and these are discussed in section 5.3.

Hammersley's discussion of his second criterion, relevance, is also highly relevant to HTA. He warned against an over-restrictive interpretation of the relevance criterion, which identifies relevance with the capacity of a piece of research to resolve the problems faced by some group of practitioners. This is further discussed in section 5.4.

5.2.4 Summary

So far, we have outlined a number of widely different positions in relation to the criteria that can and should be applied in assessing qualitative research. First, there are those who argue that any attempt to apply criteria is doomed to failure because the very idea of criteria is incongruent with the anti-realist assumptions, which are held to be central to qualitative research. We have argued that while this position is logically consist-

[33] See section 4.1.2.1.

ent, it is unnecessarily constraining and is likely to render qualitative (or indeed any research) of very limited usefulness in HTA and other applied fields. The second position is grounded in similar anti-realist assumptions. Its proponents argue that qualitative and quantitative research are situated in fundamentally different paradigms and that, therefore, conventional criteria should be replaced with ones that are more congruent with the assumptions of the qualitative paradigm. This position again raises again some difficulties for those who wish to use qualitative research to assess health technology. First, as Smith (1984) has argued, there is a fundamental contradiction in attempting to combine an anti-realist perspective with the search for criteria by which such research may be assessed. Phillips (1987) has summed up the problem succinctly: "Some qualitative researchers claim that there is no 'truth' and still want their account of X to be believed" (11).

Second, the assumption of multiple realities, upon which this position depends, raises serious questions about the possibility that health technology can be assessed in any meaningful sense at all. If we accept this position we are forced to concede that any research report must be treated as just one account among many, incommensurable accounts (Hammersley, 1992). Clavarino and co-workers (1995) criticised this position as tantamount to "methodological anarchy" and urged us to resist the "siren calls of relativism" (225). If the findings of any research cannot be taken to represent even an approximation of truth, then there one has to ask why commissioners should invest public money in funding such research.

Thankfully, both qualitative and quantitative research are perfectly consistent with a rejection of such anti-realist assumptions. The alternative of subtle realism, which Hammersley advocated (1990; 1992d; 1992e) is likely to promote the kind of research that holds most promise for research in HTA. It allows us to hold on to the search for truth as a "regulative ideal" (Phillips, 1987:23), while at the same time accepting that it is impossible to be absolutely certain that such truth has been attained. The criteria by which all research, whether qualitative or quantitative, should be assessed are those of validity and relevance. If research studies are to be used to inform the development and application of health technologies, we must, first and foremost, have some confidence that their findings are true. Likewise their relevance to the concerns identified by commissioners must be clear. While it is impossible, in either qualitative or quantitative research to provide absolute proof of the validity of one's findings, researchers have at their disposal a number of ways in which they can limit the "likelihood of error" (Hammersley, 1990). Our commitment should be to what Dewey has termed "warranted assertability" (Phillips, 1987). Given the differences between qualitative and quantitative research, in terms of the problems that each addresses and the research practices used, the means by which the likelihood of error will be limited can be expected to vary between the two research traditions. As a result, the means by which we may assess researchers' success in limiting error will also vary. In the next section, we turn to the ways in which it has been proposed that the success of qualitative researchers in limiting error should be assessed. In doing so, we are working from the position that while the criteria by which all research should be judged are identical, the means by which such judgements should be arrived at should be method-appropriate (Jensen, 1989; Glaser and Strauss, 1965a).

5.3 Assessing the validity of qualitative research

In this section, we review various means of improving the validity of qualitative research, which can be used as a basis for judging the rigour with which qualitative research has been carried out. We shall assess each in relation to their usefulness in terms of their contribution to limiting the risk that the findings of research are erroneous. Bearing in mind the impossibility of establishing the truth of any research beyond all possible doubt,[34] we shall resist the temptation to provide a rigid checklist of rules which qualitative research must follow if it is to be deemed valid. As Phillips (1987) argued, there are no methods that will yield sound data or true conclusions regularly. The formulation of truth is not "simply a matter of finding and following certain analytic procedures" (21). There can be no simple algorithmic criteria that can be applied unproblematically to judge the goodness of qualitative research (Hammersley, 1992e). Assessment of research always involves judgement and there is always potential for judgement about the application of any criteria. The risk of checklists is that they become rigid constraints which become an end in themselves rather than serving to

[34] See section 5.2.2.3.

enhance the validity of the study (Marshall, 1985). They may be more effective in producing consensus than in guaranteeing truth (Phillips, 1987). They may lead to defensive behaviour among researchers and produce sanitised but insignificant findings (Marshall, 1985). In most cases, design decisions reflect a trade-off between various means of enhancing the validity and relevance of research findings. Most research is carried out within budget constraints and researchers are forced to make judgements about the priority to be given to one or other approach to enhancing the validity and/or relevance of a study in the light of budget and practical constraints. These are decisions that are familiar to quantitative researchers and apply equally to qualitative research. As Dingwall has commented elsewhere: "One of the great methodological fallacies of the last half century in social research is the belief that science is a particular set of techniques; it is, rather, a state of mind, or attitude and the organisational conditions which allow that attitude to be expressed" (Dingwall, 1992:163).

Even if we were to accept the usefulness of such checklists, we would be left with the problem of deciding what items should appear on them. In next two sections, we consider two techniques (respondent validation and triangulation), which are sometimes proposed as validity tests, before going on to outline some of the principles upon which evaluators' judgements about the validity of research findings might more appropriately be made.

5.3.1 Respondent validation or member checking

A number of authors have recommended the use of a practice variously known as **member validation**, **host recognition** or **informant checking** as a means of establishing the validity of the findings of qualitative research (Frake, 1964; Frake, 1975; Guba, 1981; Goodwin and Goodwin, 1984a; Lincoln and Guba, 1985; Sandelowski, 1986; Guba and Lincoln, 1989; Walker, 1989; Beck, 1993). As Bloor reported, advocates of such approaches hold: "Analysis to have been validated when a correspondence is demonstrated between the investigator's description and the descriptions of members of the collectivity that is being investigated" (Bloor, 1983:156).

Such approaches include:

- validation of the researcher's taxonomies by attempting to predict members' descriptions in the field
- validation of the researcher's analysis by the attempted prediction of members' reactions to hypothetical cases
- validation of the researcher's analysis in relation to the researcher's ability to 'pass' as a member of the setting (e.g. Frake, 1964)
- validation of the researcher's analysis by asking those studied to judge the adequacy of the researcher's analysis for themselves.

The first three kinds of member validation exercises, noted above, are relatively unusual in qualitative research. As tests of validity, they each pose some problems. Both Emerson (1981) and Bloor (1983) have pointed out that, just because a researcher is able to pass as a member in a setting, does not guarantee the validity of the researcher's analysis, as members may tolerate a considerable degree of deviance in fellow members. In addition, as Emerson (1981) observed, attempts to apply the rules of behaviour, derived from research, in specific situations will encounter the same difficulties as all attempts to follow rules, insofar as such rules can never completely specify the procedures by which they should be applied in particular situations. These member validation exercises may serve as a way of identifying and reducing error. They can never provide ultimate proof of the validity of findings.

The limited use of the first three approaches to member validation is partly due to the practical difficulties in mounting such exercises. In contrast, feeding back analyses to research participants and asking them to judge their adequacy, has gained considerable popularity in recent years. Bloor summarised this approach as suggesting that 'qualitative findings are validated to the extent that collectivity members recognize and endorse the sociologist's account of the social world.' (Bloor, 1983:157). Sandelowski (1986) advocated this approach and argued that a study can be judged credible if: "It presents such faithful descriptions or interpretations of a human experience that the people having that experience would immediately recognize it, from those descriptions or interpretations, as their own" (28).

Lincoln and Guba (1985) presented respondent validation as the ultimate check on the credibility of a study's findings. Guba (1981) argued that the ultimate test of validity lies in isomorphism between a study's findings and respondents' perceptions. Writing with Lincoln, nearly a decade later, he described member checks as "the single most crucial technique for establishing credibility" (Guba and Lincoln, 1989:239). Among the functions that Lincoln and Guba identified for such member checks are the opportunity for assessing

the intent of a given action,[35] for correcting errors of fact and interpretation, for obtaining additional information, for putting the respondent on record as having agreed that the researcher 'got it right', for summarising the findings and for judging the overall adequacy of the study. They took as an example of the strength of such member checks a study carried out by Skrtic and co-workers (1985). Member checking was carried out in this study and failed to turn up a single suggestion for correction of interpretation. Guba and Lincoln (1989) concluded, on the basis of this member checking exercise, that: "No person, no matter how powerful or remote from power, at any site, felt that her or his construction had been mis-represented" (240).

Guba and Lincoln interpreted this as evidence of the "kind of trust the hermeneutic process, carried out with integrity, can engender" (240). Unfortunately, this is not the only conclusion that could be drawn from the failure of member checks to uncover any dissent. We could conclude that the failure to uncover any errors of interpretation in the member-checking exercise should give cause for concern about the power of such member checks to uncover error rather than complacency about the validity of the original research findings.

A number of authors have raised such doubts about the strength of such member checks as exercises in validation (Emerson, 1981; Bloor, 1983; Silverman, 1985; Phillips, 1987; Bryman, 1988; Emerson and Pollner, 1988; Hammersley, 1990; Hammersley, 1992d; Hammersley, 1992e; Henwood and Pidgeon, 1993; Bloor, 1997). In the first place, the concern with **isomorphism** between informant and researcher perceptions (Guba, 1981) is problematic. As both Bloor (1983) and Emerson and Pollner (1988) observed, direct comparison between researcher and member accounts is simply not feasible. Drawing on Schutz (1967), both Bloor (1983) and Emerson and Pollner (1988) argued that the accounts produced by researchers and informants will each be formulated in the light of different purposes at hand and can be expected to differ from one another in ways which have no bearing on their validity or otherwise. This means that such member checks are limited to asking members to judge whether a researcher's account represents a "legitimate elaboration and systematization of the member's account" (Bloor, 1983:157).

Even where researchers acknowledge this limitation of member checking, significant problems remain, as Bloor (1983) observed. Bloor drew upon two studies in which he attempted to use different approaches to member validation (Bloor, 1976[36]; McKeganey and Bloor, 1981; McKeganey and Bloor, 1991). Both of these were in healthcare settings. In the first, Bloor was studying regional variations in adeno-tonsillectomies in two regions of Scotland (Bloor, 1976). He observed outpatient clinics in each region, so as to be able to compare the surgical assessment practices in different regions to determine whether the variation in rates of surgery could be explained by different surgeons' decision making criteria. Having analysed his data, Bloor fed back written summaries of his findings to each surgeon, and then carried out individual interviews with each surgeon, to determine whether the surgeon endorsed his analysis of the criteria which they employed. The second study was also an observational study, this time in a psychiatric day hospital (McKeganey and Bloor, 1981; McKeganey and Bloor, 1991). The particular focus of the study was upon the relationship between the informal patient culture and the hospital's group therapy programme. Here, Bloor again produced a written summary of his analysis and fed this back to both patients and health professionals from the setting he had studied. He then carried out two group discussions, one with some of the patients and the other with the staff of the day hospital. In these group discussions he encouraged members to give their reactions to his precirculated draft. In a similar exercise, Emerson and Pollner (1988) took the findings from a study of the management of psychiatric emergencies in California, back to personnel involved in delivering such services.

Bloor (1983; 1997) and Emerson and Pollner (1988) identified a number of problems with their attempts at member checking, which cast doubt upon the feasibility of using such methods as a test of validity. First, they concluded that research participants could not be relied upon to have read the draft of his analyses with sufficient attention or with the kind of critical spirit, which is necessary for the task to be carried out successfully. Emerson and Pollner (1988) referred to this as the problem of **textual reference**. One simply cannot take for granted that members will have engaged with the

[35] See section 3.2.2.1 for a discussion of some of the difficulties associated with relying upon members' interpretations;
[36] See appendix 1 for details of this study.

materials or that they will have focused upon on the matters that are of direct concern to the researcher. While the focus of the research may be of great interest to the researcher, there is no guarantee that participants will share such enthusiasm. Emerson and Pollner reported that, in their attempted member checking in a study of psychiatric emergency teams, they were sometimes criticised by members for things which they had not written and endorsed for thing which they had not said. Similarly, it is impossible to disentangle participants' responses to the analyses from their situated behaviour within the context of the validation interview. As Emerson and Pollner argued, we must understand validation exercises as "sources of information about a **social** world and episodes situated **within** and expressive of that world" (Emerson and Pollner, 1988:194). This is a special case of the problems we have identified with using interviews as a source of objective data.[37] For example, in the adeno-tonsillectomy study, Bloor found that a surgeon who initially said that he found nothing to disagree with in the draft report, was able on persistent questioning to identify several items which he felt were questionable. It is difficult to separate out the doctor's true feelings about the report from his sense that the researcher required him to identify something, anything, which was inaccurate in the report. Emerson and Pollner (1988) referred to this as the **transactional context**, observing that the way in which the researcher frames his questions and probes members' responses may tacitly direct or pre-structure the responses received. Bloor also noted that the interviews and discussions he set up were marked by what he described as consensus-seeking behaviour by both himself and the participants. Member checks are governed by the same rules of etiquette and polite behaviour as other social encounters (Bloor, 1997). Emerson and Pollner referred to this as the **relational context** of member checks and observed that, in their study, the responses of one of those with whom they carried out member checking could be interpreted as his attempt to 'do friendship' with the researcher. The concern to avoid contentious issues and to resolve any conflict that did arise, again complicates any attempt to use member-checks as a straightforward test of validity. Another problem raised by both Bloor and by Emerson and Pollner is that one cannot assume that members will act as unbiased assessors when asked to comment upon draft reports. Rather, as members of the setting which has been studied, one would expect them to have their own agenda and that this would be reflected in their response to draft findings. Emerson and Pollner referred to this as the **organisational context** of member checks. They commented: "In divided or politicized organizational contexts, expressions of support or, for that matter hostility, are apt to be constructed in the light of the research reports consequences for promoting one or another of the competing interests" (195).

Finally, Bloor noted that informants' responses were not necessarily consistent either within the interviews, or indeed across time. Similarly, Emerson and Pollner (1988) reported that the responses that members gave to their report were sometimes ambiguous.

Whereas in the adeno-tonsillectomy study, Bloor interviewed each doctor separately in relation to the analysis of his or her own practice, in the psychiatric day hospital study, he attempted to elicit responses from both staff and patients to the same analyses. This raised further problems for the interpretation of members' responses to his findings. In particular, he found that certain aspects of his reports were endorsed by one group but rejected by the other, and vice versa. The decision to carry out group discussions rather than individual interviews arose from Bloor's dissatisfaction with the way in which he, as interviewer, had shaped the content of the interviews with the surgeons in the previous study. By setting up group discussions, he hoped that he might be able to minimise his own impact on the data obtained. However, he found that, in the absence of a strong lead from himself, one member of each of his discussion groups, acted as chair and, as a result, constrained the context in much the same way as Bloor himself had done in the previous study.

As both Bloor and Emerson and Pollner have demonstrated, member checks cannot be treated as unproblematic tests of validity. Responses from members are not "immaculately produced, but rather are shaped and constrained by the circumstances of their production" (Bloor, 1983:171). This does not, of course, mean that such exercises have no value. It is the claim that such exercises can be used to **establish** validity, to which Bloor and others have objected. Emerson (1981) suggested that such exercises simply offer "additional equivocal evidence; additional

[37] See sections 4.3.6 and 4.3.7.

situationally influenced statements must then be interpreted, just as the earlier statements were, in order to establish a version of the member's perspective" (362). Both Bloor (1983) and Emerson (1981) identified benefits in using member checking that are consistent with the aim of error reduction, which we have identified as an appropriate goal in carrying out and assessing qualitative research. These included easing access negotiations to sensitive settings, by reassuring members that they will be able to review the findings before publication, and generating material which may cause the researcher to re-visit and revise the analysis. As Bloor pointed out, members' responses to the analyses must be treated as data, not as a test of validity. Emerson (1981) suggested that such data offer an opportunity for "reflexive elaboration" (362) of the original analysis. Feedback from informants is most useful when it challenges rather than endorses the analysis. As Bloor commented:

> Ironically, but inevitably, negative member reactions may act as a spur to reanalysis, but positive member reactions cannot be taken to indicate that the analyst's task is completed. By attempting to incorporate members' caveats and criticisms into a reworked analysis the researcher may broaden his or her analysis in a manner similar to that of analytic induction. (172)

5.3.2 Triangulation

Triangulation has been proposed as a means by which the validity of qualitative research findings may be enhanced or established (Trend, 1979; LeCompte and Goetz, 1982; Goodwin and Goodwin, 1984a; Lincoln and Guba, 1985; Kirk and Miller, 1986; Mitchell, 1986; Sandelowski, 1986; Merriam, 1988; Jensen, 1989; Marshall and Rossman, 1989; Stange and Zyzanski, 1989; Walker, 1989; Hammersley, 1990; West, 1990; Flick, 1992; Hammersley, 1992f; Steckler *et al,* 1992; Beck, 1993; Cowman, 1993; LeCompte and Preissle, 1993d; Greene, 1994; Miller and Crabtree, 1994; Morse, 1994; Dootson, 1995; Keen and Packwood, 1995; Janesick, 1994). The concept of triangulation is drawn from military, navigational or surveying contexts (Jick, 1979; Flick, 1992; Hammersley and Atkinson, 1995; Nolan and Behi, 1995; Janesick, 1994). Hammersley and Atkinson (1995) describe the way in which triangulation is carried out in surveying:

> For someone wanting to locate their position on a map, a single landmark can only provide the information that they are situated somewhere along a line in a particular direction from that landmark. With two landmarks, however, one's exact position can be pinpointed by taking bearings on both; one is at the point on the map where the two lines cross. (231)

Campbell and Fiske (1959) are credited with having introduced the concept of triangulation into social research. They argued that:

> When a hypothesis can survive the confrontation of a series of complementary methods of testing, it contains a degree of validity unattainable by one tested within the more constricted framework of a single method ... Findings from this latter approach must always be subject to the suspicion that they are method-bound. (82)

Campbell and Fiske advocated **multiple operationisms** in the measurement of concepts, as a means of overcoming the biases which attach to a single method. To the extent that the findings from a range of measures converged, they argued that one could conclude that method specific biases had been overcome. Webb and co-workers (1981) argued that such triangulation represents the most persuasive evidence of the validity of a proposition:

> Once a proposition has been confirmed by two or more independent measurement processes, the uncertainty of its interpretation is greatly reduced. ... If a proposition can survive the onslaught of a series of imperfect measures, all with their irrelevant error, confidence should be placed in it. (35)

Whereas Campbell and Fiske sought to use triangulation to confirm the validity of measurements of discrete constructs, Denzin (1970) extended the concept of triangulation to the confirmation of the accuracy of the findings of a research project, more generally. Denzin defines triangulation as "the combination of methodologies in the study of the same phenomenon" (Denzin, 1978; Krefting, 1991). Denzin (1970) identified four types of triangulation.

- **Method:** different methods are used to address the same phenomenon.
- **Data:** different data sources are used to study the same phenomenon.
- **Investigator:** different investigators are used in the same study.
- **Theoretical:** different theoretical models are used in the same study.

More recently, other authors have added to the list. Burns and Grove (1993) proposed **analysis triangulation** in which the same data set is analysed using two or more techniques. Janesick (1994) has suggested **interdisciplinary triangulation**, in which different disciplinary perspectives and models should be brought to bear upon a single study.

While most authors agree that the use of multiple methods in qualitative research is

to be recommended, a number have expressed concern that it should not be used slavishly (Stake, 1994), as an end in itself, or seen as "an inherent good" (Knafl and Breitmayer, 1991). Knafl and Breitmayer (1991) argued that if triangulation is used inappropriately it may compound the weaknesses of a research project rather than strengthen it.

One of the problems surrounding the concept of triangulation is the ambiguity about what the purpose of triangulation is. The lack of consensus about the meaning of the term has been recognised by Nolan and Behi (1995) and Knafl and Breitmayer (1991). Arguments in favour of triangulation can be divided into two categories.

- Those that propose the use of multiple data sources/methods/investigators/theories, as a means of testing the validity of the findings of one of the approaches used. This is the original sense of the concept as employed by Campbell and Fiske (1959), Webb and co-workers (1981), Denzin (1970), and Goodwin and Goodwin (1984a). The emphasis is upon counterbalancing the distorting effects of any single approach and the aim is to establish the convergent validity of findings drawn from complementary approaches. LeCompte and Preissle (1993d) argued that triangulation permits data collected in one way to be used to cross-check the accuracy of data collected in another.
- Those that propose that multiple methods should be used in the interests of completeness (Geertz, 1973; Jick, 1979; Silverman, 1985; Kirk and Miller, 1986; Duffy, 1987; Mathison, 1988; Oiler Boyd, 1993a; Morse, 1994). Jick (1979), for example, argued that multiple methods may "uncover some unique variance which otherwise may have been neglected by single methods" (138). Oiler Boyd (1993a) argued that triangulation increases the comprehensiveness of research by providing a more complete understanding of the phenomenon than could be achieved through a single method. Similarly Knafl and Breitmayer (1991) drew on their study of children with chronic illness to demonstrate the importance of multiple methods for the comprehensiveness of their findings, whereas they see triangulation as convergent validation as more appropriate for studies which are directed at the measurement of a single variable. Silverman (1993), who is otherwise highly critical of triangulation, conceded that the use of multiple sources is helpful in overcoming the partiality of data drawn from a single source (157).

5.3.2.1 Triangulation as a means of extending the comprehensiveness of a research study

There can be few objections to the use of multiple methods to extend the comprehensiveness of the findings of qualitative research. Indeed, the search for negative evidence[38] and fair dealing in research[39] are likely to be enhanced where data is sought from multiple sources, using multiple methods. One of important contributions of multiple methods is that they may encourage us to pay attention to the different perspectives which may be held and the way in which particular research accounts are inevitably the product of the circumstances in which they are produced. Dingwall, (personal communication) cited in Silverman (1993), suggested that triangulation may be useful where it reveals differences in the kinds of account of an agency's work which are revealed in different contexts. Jick (1979) argued that whereas, when triangulation is treated as a test of validity the concern is to establish the convergence of data from different sources, when one is concerned with achieving comprehensiveness, divergent findings from two or more methods are often the most useful. Where data from two sources or methods yield dissimilar results, the researcher is encouraged to consider how such differences have arisen and this offers an opportunity for enriching the analysis. Trend (1979) argued that researchers should exploit the opportunities, offered by triangulation, for allowing different viewpoints to arise and mature, rather than 'strangling them at birth' in the interests of consensus. Similarly, Hammersley and Atkinson (1995) suggested that differences between sets or types of data may just as important as similarities. Bryman (1988) made a similar point when he suggested that, when the use of multiple methods throws up discrepancies, these discrepancies should be made the topic of investigation since they may well lead to a more sophisticated understanding of the phenomenon under study. Thus, as Janesick (1994) suggested, multiple methods may become an **heuristic tool** (215).

Both Fielding and Fielding (1986) and Denzin and Lincoln (1994b) have suggested that, from this perspective, triangulation should be considered, not as a test of validity, but as an alternative to it. Triangulation should be used to reveal the varied dimensions of a phenomenon, not in the expectation that different data sources will confirm one

[38] See section 5.3.6; [39] See section 5.3.7.

another unproblematically. Denzin and Lincoln (1994b) summarised this point: "The combination of multiple methods, empirical strands, perspectives and observers in a single study is best understood then as a strategy that adds rigor, breadth and depth to any investigation" (2).

5.3.2.2 Triangulation as a test of validity

Bloor (1997) summarised this alternative version of triangulation in which: "Findings may be judged valid when different and contrasting methods of data collection yield identical results on the same research subjects" (38).

This argument for triangulation as validation rests upon the assumption that the weaknesses in a given method will be compensated for by the strengths of another. This involves the further assumption that multiple measures do not share the same weaknesses (Jick, 1979). Smith (1975) suggested that triangulation "finds its main value in disconfirming the tenability of arguments that findings are artifacts of particular methods" (290). It is in this version of triangulation that the origins of the term in surveying and navigation are most evident. The argument is that by using multiple methods researchers should be able to get a 'better fix' on the phenomenon of interest (Denzin and Lincoln, 1994b:2), just as in surveying the use of more than one reference point allows the position of an object of interest to be located.

While the usefulness of multiple methods for increasing the comprehensiveness of qualitative research is generally accepted, the idea that triangulation can be used to validate research findings has been subject to criticism from a number of quarters. On the one hand, predictably perhaps, it has been criticised by some of those who operate from a radical relativist (e.g. Smith, 1984) or postmodernist (e.g. Richardson, 1994) position. Their objection to triangulation as validation is that it assumes a single fixed point upon which data collected from different sources or using different methods can converge. This is clearly problematic for those who assume that there are multiple realities and that there is no objective or true reality to be confirmed.[40]

However, triangulation as validation has also been criticised from other quarters, by those who would distance themselves from such relativist assumptions. A number of objections, both philosophical and practical, have been raised against the use of triangulation as a test of validity.

First, Silverman (1985; 1993) has suggested that the fundamental problem with triangulation is that, by definition, it seeks to overcome the context-boundedness of data. The reason for juxtaposing materials gathered using different methods is to allow the researcher to extract the data from the circumstances of their production. However, as Silverman argued, this leaves us unable to analyse their sense in context. Such analysis in context is at the very heart of qualitative research.[41]

Second, the pursuit of validity through triangulation runs the risk of persuading researchers to focus their analyses on the search for a single master reality or objective truth, when their efforts might be better directed towards uncovering the situated work which different versions do in different contexts (Dingwall, 1981). The limitations of using data from one source to validate those from another, are well illustrated by the study that Stimson and Webb (1975)[42] carried out in general practices in South Wales. Stimson and Webb combined participant observation in general practice consultations with interviews with patients. They were struck by the inconsistencies between the reports which patients made about their interactions with doctors and those which they actually observed taking place. If the authors had been committed to a conventional approach to triangulation as validation they would have been obliged to adjudicate between their observations and the accounts given by patients, presumably judging patient accounts as 'invalid' in the light of their observations. However, they chose, instead, to treat the interviews as situated accounts, which offered respondents an opportunity to redress the power balance of doctor–patient relationship.

Silverman (1993) drew on Garfinkel's discussion of ironies (Garfinkel, 1967) to argue that we should not treat behaviour and talk in one setting as undercutting what appears to happen in another setting. What Silverman objects to, in triangulation, is the use of data from different sources to adjudicate between accounts. He accepted that multiple methods may in fact help to point up the situated character of accounts and are therefore to be welcomed, as long as the intention is not to judge between them (Silverman, 1993:158).

While Hammersley (1990) argued that triangulation may be used as a means of making a "more effective assessment of the likely truth of a claim" (84), writing with Atkinson, he pointed out that

[40] See section 3.2.2.1; [41] See section 3.2.2.3; [42] See appendix 1 for details of this study.

triangulation cannot be treated as a simple test of validity:

> Even if the results tally, this provides no guarantee that the inferences involved are correct. It may be that all the inferences are invalid, that as a result of systematic or even random error, they lead to the same, incorrect conclusion. What is involved in triangulation is not the combination of different kinds of data per se, but rather an attempt to relate different sorts of data in such a way as to counteract various possible threats to the validity of our analysis. (Hammersley and Atkinson, 1995:231–2.)

Similarly, Bloor (1997) suggested that triangulation may yield data which is relevant to validation but cannot be seen as constitutive of validation. Bloor identified a number of problems with the use of triangulation to test the validity of findings. First, he argued that, if we assume that, given a particular research topic, there will be one best method of investigating it, triangulation will necessarily involve juxtaposing findings from a superior method with those from an inferior method. This will not pose any problems if the findings from the two methods converge. However, if they diverge, should one set aside the findings from the superior method because they are not supported by the findings from the inferior method? Or would one conclude that the differences arise from the shortcomings of the inferior method? As Bloor pointed out, triangulation cannot be treated as a test of validity if it can only corroborate findings and never refute them (see also Trend, 1979).

Second, Bloor pointed to some practical difficulties in comparing data collected using different methods. He drew on his study of death certification practices to illustrate this point. In this study he combined in-depth interviews with clinicians who were frequently involved in death certification, with an exercise in which he asked the same clinicians to fill out dummy death certificates on the basis of detailed case summaries which he provided. He found that, in the interviews, clinicians described their practices in very general terms, whereas the dummy certificate exercise required them to deal with very specific instances. While the data he collected from the two techniques were superficially similar, he could not be sure that this was not an artefact of the lack of specificity in the interviews, compared with the dummy cases. Similarly, where there were discrepancies, these could be explained by the defeasibility of the general rules offered in the interviews which were always subject to qualification in the light of particular circumstances. Jick (1979) made the same point when he argues that, 'it is a delicate exercise to decide whether or not results have converged'. (607)

There are then very significant problems with adopting triangulation as a test of the validity of research findings. This is not to discredit the usefulness of using a range of methods or sources of data when investigating a phenomenon. Where such methods are chosen on sound theoretical grounds, rather than seen as an 'inherent good' they may add to the comprehensiveness of a study and stimulate reflexive analysis[43] of the data from different sources in relation to the circumstances of their production. In these circumstances the decision about whether to use a single method or multiple methods will often be guided by resource constraints. As Bryman (1988) pointed out triangulation is costly in terms of both time and money (see also Jick, 1979; Mitchell, 1986; Morse, 1994; Dootson, 1995). It also demands a skill mix from researchers which, given the specialised nature of much graduate research training, may be difficult to find.

Having discussed two potential candidates for a validity checklist, and found them both wanting, we now move on to discuss five principles in relation to which judgements about the validity of research findings may be made.

5.3.3 Clear exposition of data collection method

One of the strengths of qualitative research is the recognition that data must always be understood in relation to the context of their production. Qualitative researchers recognise that research activity "**inevitably** shapes and constitutes the object of inquiry" (Henwood and Pidgeon, 1993:106 original emphasis). It is therefore incumbent upon the researcher to create a record of the research process, which "can stand independently of the observer" (Dingwall, 1992:169). This means, among other things, that the researcher must describe, in detail, the process by which the data on which the analysis is based were collected (Glaser and Strauss, 1965a; Glaser and Strauss, 1967; Schatzman and Strauss, 1973; Guba, 1981; LeCompte and Goetz, 1982; Athens, 1984; Marshall, 1985; Jensen, 1989; Silverman, 1989; Beck, 1993; Henwood and Pidgeon, 1993; Altheide and Johnson, 1994; Olesen, 1994). Such descriptions

[43] See section 5.3.5.

of the circumstances of the research enable readers and researchers to exercise what Glaser and Strauss (1965a) have described as their joint responsibility to judge the evidence upon which claims derived from research are based. Such judgements will take account of the process by which the researchers arrived at their conclusions, including, 'the range of events the researcher saw, whom he interviewed, who talked to him, what kinds of experiences he had, and how he might have appeared to various people whom he studied (9). Athens (1984) argued that: "Scientific credibility is not an ascribed quality of a study, but rather an achieved one. Thus, a study is neither intrinsically credible nor incredible; rather a researcher must make it so" (265).

An important part of establishing the credibility of the findings of a research study lies in giving an adequate account of the circumstances of their production. Athens identified two particular aspects of data collection that should be attended to in research reports. First, the report should include details of how the researcher gained access to the persons, groups or organisations being studied, and how such access was maintained or increased during the project. Formal permission to carry out research in a particular organisation does not necessarily bring with it access to all settings within that organisation, and the data which the researcher obtains will be limited to that which (s)he is permitted to see or hear within each setting. Similarly, in interviews, the information made available to the researcher will be limited to that which informants are willing to share with the interviewer. Where researchers spend a prolonged period in the field, the extent and quality of their access may be expected to change over time. This again needs to be documented to enable the reader to judge the credibility of research findings in relation to the circumstances of their production.

Second, Athens argued, the researcher who wishes to establish the credibility of his or her findings, is required to describe in detail the means by which the data were produced. Taking the example of in-depth interviews, he insisted that the researcher must describe in detail how the interviews were conducted. This would include descriptions of the nature of the questions asked and the circumstances under which the interviews were carried out.

Similarly, Altheide and Johnson (1994) suggested that there are a minimal set of issues which are encountered in almost all studies and which must be detailed in research reports, in order to allow the reader to assess the degree of likely error in any research study. These include negotiation of access, self-presentation of researcher, trust, the researcher's role within the setting, any mistakes made by the researcher, the types of data collected and how such data were collected and recorded.

More recently, several authors (e.g. Jensen, 1989; Howe and Eisenhart, 1990) have related the need for clear specification of the circumstances in which data were collected to conventional concerns with reliability. They have suggested that researchers should present their methods so clearly that other researchers should be able to use the original report as an operating manual by which to replicate the original study. Such an operating manual should include, at a minimum, details of the researcher's role within the setting, the selection of informants within the setting, and the social conditions within which data were collected.

5.3.4 Clear exposition of process of data analysis

In recent years, a number of authors have criticised qualitative researchers for their failure to clarify the process by which their findings have been derived from the data collected (Guba, 1981; Lincoln and Guba, 1985; Silverman, 1985; Sandelowski, 1986; Guba and Lincoln, 1989; Jensen, 1989; Mechanic, 1989; Dingwall, 1992; Beck, 1993; Henwood and Pidgeon, 1993; Altheide and Johnson, 1994). They have called for data analysis to be made "semi-public and not magical" (Silverman, 1989) and "public and reproducible" (Dingwall, 1992). This concern relates both to the process by which data are coded and categorised in qualitative research and to the way in which conclusions are drawn from the data that have been collected.

The adequacy of any analysis depends upon the nature and quality of the process that is used to organise and interpret the data upon which it is based. Whereas in quantitative research, the researcher is usually required to operationalise the concepts of interest, before the data are collected, the inductive nature of much qualitative work[44] means that this is often not appropriate. However, this does not relieve the researcher of the obligation to develop clear definitions of the concepts and categories, which are developed in the course of the research, so that their meaning

[44] See section 3.2.1.2.

and application is unequivocal. Hammersley (1990) suggested that the researcher must give attention to two aspects of each of the key concepts used in presenting research findings: intension (the concept's relationship to other concepts) and extension (its relationship to instances) (76). In considering a concept's **intension**, we are concerned with the network of concepts to which any particular concept belongs. It is from its place in such a network that any particular concept gets its meaning. A concept's **extension**, on the other hand, concerns what would or would not count as an instance of a given concept. Once again, it is incumbent upon the researcher who wishes to establish the credibility of his or her conclusions, to demonstrate that key concepts have been defined unambiguously and coherently.

The clarification of concepts is often ongoing throughout a research project, as the researcher exposes the data to the rudimentary classification system and the classification system is refined in the light of the data. It is important, however, that by the time the results of the study are presented, a clear definition of each concept should have been arrived at, which is publicly communicable and defensible. Some authors have recommended the use of inter-rater reliability checks as a means of clarifying concepts and ensuring their consistent application to the data (Goodwin and Goodwin, 1984a; Atwood and Hinds, 1986; Jensen, 1989; Hinds *et al*, 1990; Beck, 1993; Seidel and Kelle, 1995). For example, Hinds and co-workers (1990) suggested assessing inter-rater reliability of coding by asking a panel to re-sort randomly selected data units, using the coding criteria previously developed and applied by the research team. By calculating the level of agreement between the coding panel and the researcher team, it is possible to assess the explicitness of the definitions of the concepts developed in the study.

Resource constraints may make the use of such inter-rater reliability exercises impractical for many research studies. However, the credibility of all research findings will be enhanced wherever the researcher provides comprehensive definitions, which make it clear why the phenomena of interest have been labelled in certain ways. Such definitions are a public product which make the researcher's tacit assumptions explicit and allow the reader to evaluate findings in relation to the definitions employed (Henwood and Pidgeon, 1992).

Similarly, the researcher must demonstrate that the conclusions that (s)he has drawn are justified in relation to the data collected. The first requirement here is that the data on which the analysis is based should themselves be trustworthy. A number of ways in which such trustworthiness can be enhanced are discussed in the literature. The use of mechanical recording (video or audio recording) wherever possible is recommended (Jensen, 1989; Waitzkin, 1990; Beck, 1993; LeCompte and Preissle, 1993c). The use of standardised rules for transcribing data is also recommended (Waitzkin, 1990) as a means of limiting the risk that data is misinterpreted at the analysis stage. A number of authors emphasised the importance of presenting low-inference descriptors such as verbatim accounts and concrete and precise descriptions (Beck, 1993; LeCompte and Preissle, 1993c).[45] Hammersley (1990) suggested that in assessing the trustworthiness of research data we should ask some commonsense questions including whether it is likely that the informant would have had access to the events (s)he describes, whether (s)he might have ulterior motives for misleading the researcher and so on. Similarly, confidence in the data on which analyses are based will be strengthened where the researcher has spent an extended period in the setting which (s)he has studied (Glaser and Strauss, 1965a; Guba, 1981; LeCompte and Goetz, 1982; Merriam, 1988; Guba and Lincoln, 1989; Jensen, 1989; LeCompte and Preissle, 1993c). Such prolonged engagement in the field allows the researcher to search thoroughly for cases which augment, disconfirm or corroborate their early conclusions (LeCompte and Preissle, 1993c) and to identify ways in which the researcher's presence has contributed to the data obtained.

[45] It should not be assumed that the notion of a verbatim account is itself unproblematic (Graffam Walker 1986; Green *et al,* 1997). Although there are a small number of studies showing a reasonable measure of inter-transcriber reliability and fidelity to the original (Lindsay and O'Connell 1995; Patterson *et al,* 1996), the transcript is the product of the particular notation system adopted, particularly where markers for non-semantic features (pauses, overlaps, intonation, etc.) are introduced. As Sandelowski (1994) points out, the notion of a verbatim transcript depends upon a strongly realist assumption that it is possible to produce a literal description of any event in the world. As for observations or interviews, transcripts are means of selecting from and organising the researcher's experience in terms that appear to be significant and useful. In evidential terms, transcripts may be an improvement on the recording of a researcher's subjective impressions or feelings, at least by the criteria of accessibility to the original data or event but they are not a standard of pure gold.

A number of authors have emphasised the importance of displaying enough data to allow the reader to assess whether the interpretations are adequately supported by the data (Athens, 1984; Marshall, 1985; Silverman, 1989; Hammersley, 1990; Waitzkin, 1990; Dingwall, 1992; Beck, 1993; Secker *et al,* 1995). Athens (1984) has emphasised the importance of the empirical grounding of research findings. Readers must be enabled to scrutinise the empirical observations upon which conclusions are based. To allow this, two conditions must be met. First, each major concept must be accompanied by at least some of the empirical cases which led to its development and, second, these cases must be fully presented rather than in "small, carved down slices" (264). It must be possible for the reader to assess whether the data presented by the researcher justify the claims that are being made about them (Hammersley, 1990). Such assessment depends crucially upon the extent to which the researchers have separated out the data and the analysis of that data, in presenting their conclusions (Athens, 1984; Dingwall, 1992).

The trustworthiness of researchers' analyses of their data is enhanced where researchers can demonstrate that they have considered alternative plausible explanations of their data (Marshall, 1985; Hammersley, 1990; Waitzkin, 1990). Marshall (1985) argued that researchers should lay out all possible competing interpretations of their data and present the reader with the evidence which supports the choice of these interpretations which has been made by the analyst.

Once again, some authors have proposed some formal procedures by which, where resources permit, the credibility of analyses may be examined. A number have recommended the use of peer-review panels, charged with examining the credibility of the claims, which researchers have made in the light of the data available (Goodwin and Goodwin, 1984a; Atwood and Hinds, 1986; Merriam, 1988; Guba and Lincoln, 1989; Jensen, 1989). The members of such panels would be expected to track the data to their sources and examine the coherence of the conclusions which the analysts have drawn from their data. The feasibility of such procedures will again depend upon the availability of resources.

The recognition of the importance of examining and displaying the processes by which data were collected and analyses were arrived at has been associated with the emergence of auditability as a criterion against which it is proposed that qualitative research should be assessed (Guba, 1981; Lincoln and Guba, 1985; Sandelowski, 1986; Merriam, 1988; Guba and Lincoln, 1989; Mechanic, 1989; Beck, 1993; Henwood and Pidgeon, 1993). Researchers are encouraged to create an audit trail (Guba, 1981; Lincoln and Guba, 1985; Mechanic, 1989), decision trail (Sandelowski, 1986) or paper trail (Henwood and Pidgeon, 1993). Such a trail should provide the raw materials which would allow an external auditor to determine the trustworthiness of the data on which analyses are based and the interpretive processes by which analyses have been arrived at.

5.3.5 Reflexivity

The importance of relating qualitative research data to the circumstances of their production also points up the role of reflexivity in qualitative research (Silverman, 1989; Hammersley and Atkinson, 1995). Whereas within the quantitative tradition, the emphasis is upon eliminating the impact of the researcher upon the research findings, primarily through standardisation of procedures, qualitative researchers acknowledge that there is "no way in which we can escape the social world in order to study it" (Hammersley and Atkinson, 1995:17). By reflexivity we mean sensitivity to the ways in which the researcher's presence in the research setting has contributed to the data collected and their own *a priori* assumptions have shaped the data analysis. Qualitative research calls for a level of self-conscious reflection upon the ways in which the findings of research are inevitably shaped by the research process itself and an analysis which takes such factors into account.

As Henwood and Pidgeon (1993) have observed research activity inevitably shapes and constitutes the object of its inquiry. The data which are obtained in any setting or interview situation reflect the social relationships within which they are embedded. This means that, in drawing conclusions from data, the researcher must reflect upon his or her own impact upon the setting (Altheide and Johnson, 1994) and use whatever means are available to assess the likely impact of the researcher's presence on the data obtained. There are various means of making such assessments. These include comparing the statements that are made to the researcher alone with those that are made to others in everyday situations (Silverman, 1989) and monitoring changes in the data obtained over time (Guba and Lincoln, 1989). Hammersley and Atkinson (1995) recommended that researchers should actively compare data in which the level and direction of reactivity vary.

A number of authors have recommended that the self-conscious monitoring of the researcher's impact upon the setting should be carried out alongside the data collection. Researchers are encouraged to take time out away from the research setting to reflect upon the impact of their presence upon the research (LeCompte and Goetz, 1982). The use of peer de-briefing sessions, which give the researcher an opportunity to discuss the circumstances of the data collection with a colleague who is not closely involved in the day-to-day research is recommended (Guba, 1981; Lincoln and Guba, 1985; Guba and Lincoln, 1989).

The credibility of research findings is also enhanced where researchers make the personal and theoretical biases which they bring to the research explicit in the research report (Marshall, 1985). As Hammersley (1992) argued, the assumption that there is one true version of truth, which the researcher seeks to reproduce in his or her research, fails to take account of the fact that empirical phenomena are **descriptively inexhaustible**. Since it is possible to provide multiple, true descriptions of any phenomena, we are forced to recognise the role of values and *a priori* assumptions in shaping any research account (Hammersley, 1990). Rather than treating research findings as reproducing reality, we must recognise that they are a representation of reality, which inevitably reflects the assumptions that researchers bring to their analyses (Hammersley, 1990). The implication of this observation is that the researcher must make explicit the personal and theoretical assumptions which underpin his or her work (Guba, 1981; Marshall, 1985; Merriam, 1988; Dingwall, 1992; Altheide and Johnson, 1994). Such assumptions are not necessarily static. They may well undergo modification in the course of the research as the researcher engages with the setting. It is important that such changing perspectives are monitored, documented and dealt with in the research report. Making assumptions explicit in this way, allows the reader to assess the researchers' findings in the light of the assumptions which have shaped them.

The recognition and explication of the researcher's cultural assumptions also plays an important part in guarding against perceptual bias arising from ethnocentrism (Marshall, 1985; Hammersley, 1992d; LeCompte and Preissle, 1993c).[46] Unless researchers recognise and challenge their own taken-for-granted assumptions, there is a serious danger that they will fail to recognise the significance of data that cut across those assumptions. This is a particular problem in health-related research, where the researchers' assumptions reflect the dominance of the so-called **medical model**. A failure to recognise that informants may be drawing on a different, but equally logical set of assumptions, may blind the researcher to important data.

Researchers should also, as Becker (1967a) argued, make it clear whose perspective has been adopted in a particular study. Where, for example, the research has focused upon the patients' perspective on the doctor–patient interaction, it is incumbent upon the researcher to make it clear that the patient's perspective does not represent 'reality' in some ultimate or objective way: "We warn people, thus, that our study tells us only how things look from one vantage point" (Becker, 1967a:247).

It is not only the researchers' assumptions which constrain the data that are collected and the interpretations which are made. The researcher's structural position and ascribed characteristics will also influence the research process and need to be handled reflexively in the analysis (Silverman, 1993). In recent years, particular attention has been paid to the influence of the researcher's gender on the data obtained in interviews and fieldwork. Informants have been shown to say different things to male and female researchers and individuals, access to particular parts of research settings or certain kinds of information within that setting are also related to gender (Becker, 1967a; Warren, 1988). However, as Silverman (1993) and McKeganey and Bloor (1991) have pointed out, variables other than gender, such as age and social class, may also have an important impact on the data that is available in fieldwork settings.[47] In the context of this review it is also important to note that, where the researcher is known to be a health professional, it is likely that the information which is given will reflect that which it is deemed appropriate to give to a health professional. Such information is not necessarily better or worse than that which will be given to other kinds of researchers, but it can be expected to be systematically different. Such differences need to be taken into account in presenting the findings of the research.

Another threat to the credibility of research findings lies in the risk of the researcher going native (LeCompte and Goetz, 1982; Marshall, 1985;

[46] See section 4.2.4; [47] See section 4.2.2.

Sandelowski, 1986; Beck, 1993).[48] Hammersley and Atkinson (1995) define this as "the danger of identifying with ... members' perspectives, and hence failing to treat these as problematic" (111).

Going native is recognised as a particular danger where the researcher spends a prolonged period in the research setting and comes, often without realising it, to share the assumptions of those (s)he is studying. The parallel problem of being native may arise where, as for example, in studies of medical decision making by doctors, where the researcher so shares the assumptions of the group being studied that (s)he is unable to gain sufficient distance to treat these assumptions as problematic. Recommendations for avoiding such over-rapport include regular withdrawal from the field to allow time for reflection and distancing (Guba, 1981; LeCompte and Goetz, 1982) and conscious reflexivity (Guba, 1981; Sandelowski, 1986; Hammersley, 1992d), using, where appropriate, peer de-briefing (Guba, 1981; Lincoln and Guba, 1985; Guba and Lincoln, 1989).

Researchers' claims to credibility will be strengthened where they can demonstrate that both data collection and analysis have been marked by a commitment to reflexivity and that, in arriving at their conclusions they have taken account of the circumstances of their production.

5.3.6 Attention to negative cases

Sandelowski has identified **holistic bias** as a major threat to the validity of qualitative research. She defined holistic bias as making the data look more patterned than they are. The recommended antidote to such holistic bias is the conscientious search for and presentation of cases that are inconsistent with the emerging analysis (Glaser and Strauss, 1965a; Schatzman and Strauss, 1973; Athens, 1984; Lincoln and Guba, 1985; Marshall, 1985; Silverman, 1985; Sandelowski, 1986; Phillips, 1987; Guba and Lincoln, 1989; Mechanic, 1989; Silverman, 1989; Waitzkin, 1990; Dingwall, 1992; Beck, 1993; Henwood and Pidgeon, 1993; LeCompte and Preissle, 1993c; Secker *et al,* 1995). Theoretical sampling[49] facilitates the search for negative cases (Glaser and Strauss, 1965a; Guba, 1981; Goodwin and Goodwin, 1984a; Lincoln and Guba, 1985; Hammersley, 1990; Dingwall, 1992; Henwood and Pidgeon, 1993) and encourages researchers to establish the limits of their findings (Glaser and Strauss, 1965a), through the systematic search for disconfirming cases (Dingwall, 1992).[50]

The careful study of deviant cases allows the researchers to refine their analyses (Silverman, 1989), until they are capable of incorporating all available data (Secker *et al,* 1995). The credibility of research findings is strengthened where researchers display negative cases in their reports (Silverman, 1989) and then show how the analysis can explain such apparent inconsistencies (Secker *et al,* 1995). Conceptually dense explanations are those which are capable of integrating such negative cases (Henwood and Pidgeon, 1993). Some authors (Lincoln and Guba, 1985; Guba and Lincoln, 1989) have argued that refining the analysis until it accounts for all known cases is too rigid a goal. Researchers should aim instead for an analysis which accounts for most of the available data. Others have argued that analyses should account for **all** the data (Secker *et al,* 1995). In either case, the emphasis is upon the systematic coding and analysis of all the data (Silverman, 1985), the conscientious presentation of all negative or exceptional instances (Silverman, 1989) and the importance of modifying the analysis in the light of contrary evidence. As in all science, it is the careful search for falsifying evidence which adds weight to the truth claims of qualitative research. While, as Popper has observed, such an approach can never guarantee truth, it does support the elimination of error (Phillips, 1987).

Silverman (1985; 1993) took the example of a research study of doctor–patient interactions, reported by Waitzkin (1979), to demonstrate the perils of failing to take account of negative evidence. Waitzkin's starting point was what Silverman described as a "mechanistic version of Marxism ... which reduces medicine simply to an ideological apparatus of the capitalist state" (185). Waitzkin (1979) used his data to illustrate three assertions about doctor–patient interactions which are derived from his preconceived theory.

- Doctor–patient interaction parallels the situation in the workplace where information is concentrated in the hands of the few.
- The doctor "is in a position to enforce industrial or home discipline through ideological messages about the work ethic" (604–5).
- "Objectification in doctor–patient interaction ... using machine like references for parts of the body ... parallels the use of the wage-earner or home maintainer as machine-like human capital" (605).

[48] See section 4.2.4; [49] See section 4.1.1.4; [50] See section 4.6.1.3.

While Waitzkin produced examples which seem to support these assertions, he failed to track down or follow up contrary evidence which might challenge or at least modify his conclusions. For example, he failed to subject his assertion that the use of machine metaphors in the consultation is a feature of the doctor–patient relationship within capitalism to serious test. He might for example, have asked whether such metaphors are also used in non-capitalist, industrialised countries. Similarly, he failed to integrate his findings that women receive more information than men within consultations and that doctors from working-class backgrounds give less information than doctors from upper-class backgrounds, into his analysis. Silverman (1985) summarised the problems, which are associated with this failure to take seriously findings that challenge the analyses: "In the same way as the Bible advises 'look and ye shall find', so the partisan looks and inevitably finds examples which can be used to support his theory" (Silverman, 1985;187).

By contrast, Silverman's own study of paediatric cardiac oncology clinics illustrated the way in which sustained attention to negative or discrepant cases can lead to a more sophisticated and coherent analysis that might otherwise be the case (Silverman, 1981).[51] Silverman's particular interest was in how disposals were organised and announced within the clinics. He observed that, at an initial outpatients' clinic, doctors would not normally announce the discovery of a major heart abnormality and the necessity for life-threatening surgery to parents. Rather they adopted a step-by-step method of information giving. At the first consultation they suggested the need for more tests and merely hinted at the possibility that major surgery might be required. However, there was a small number of cases where this step-by-step procedure was not employed. Rather than ignoring these cases, Silverman analysed them in some detail, seeking to understand why they differed from the other cases studied. He identified two circumstances under which the step by step procedure was abandoned and the doctor routinely presented all the available information at the first consultation. The first of these was where the child was found not to have a serious heart defect. In such cases the doctors fully disclosed the available information. The second circumstance was where, in addition to a heart defect, the child also had Down's syndrome. In these cases, the doctor's behaviour was systematically different from that observed in consultations with non-Down's syndrome children. The consideration of these negative cases allowed Silverman to develop an analysis of disposals and their announcement in paediatric cardiology clinics, which did justice to all of the data.

Similarly, Strong's analysis of the small number of cases which did not meet the rules, which he had induced from his study of doctor–patient interactions in paediatric clinics, proved to be an important analytic lever. He explained how he made use of these limiting or negative cases: "The scope of the rules has been shown first by a consideration of their outer edges, the limiting cases in which they do not apply ... one analytic strategy in these cases is to show that, although these are counter examples, they are produced by very special circumstances and occur only within these" (Strong, 1979a:233).

This discussion of the role of negative evidence in qualitative research points to one of the shortcomings of much of the qualitative research as it is currently practised. As Silverman (1989) has demonstrated, by auditing a set of issues of two of the most prestigious journals which report health-related social science research, there are relatively few attempts to build cumulative knowledge.[52] Hammersley (1985) has noted a similar deficiency in the field of education. This failure to link new work to the cumulative body of knowledge that already exists is all the more problematic given the difficulty of funding large-scale, multi-site studies. This means that the search for negative evidence, through systematic, theoretical sampling, within an individual study, is inevitably constrained. If we are to approach **practical certainty** (Gibson, 1978), by testing the limits of tentative interpretations and conclusions in a range of empirical situations, we shall have to make use of opportunities to develop earlier work in later studies. Qualitative researchers have been curiously loathe to do so. Partly this relates to the concern, at the heart of much writing on qualitative method, to avoid opportunistically imposing "theories that have dubious fit and working capacity"(Glaser and Strauss, 1967:4) upon research data. Similarly, the emphasis on the role of qualitative research in theory generation, as opposed to theory testing (Glaser and Strauss, 1967; Marshall, 1985; Strauss and Corbin, 1990; Henwood and Pidgeon, 1993) has tended to be used as a justification for stand-alone studies that make little contribution to our cumulative knowledge of the issues they raise.

[51] See appendix 1 for details of this study; [52] See section 3.2.1.2.

The tendency for qualitative researchers to limit themselves to 'one-off' studies, which are "conceived and executed in magnificent isolation" (Dingwall, 1992:171), limits the search for negative evidence which is crucial to the logic of qualitative research design. As Mitchell (1983) has pointed out the justification for the study of a single case or setting is that it has been chosen to extend current theory or to probe the plausibility of what we currently think that we know. If it is not, as Dingwall has observed, "it has little more than anecdotal value to a policy maker trying to understand how an organisation works" (Dingwall, 1992:171).

It can therefore be argued that one of the criteria by which qualitative research should be judged is the extent to which the researchers have built upon previous knowledge in their work and their success in connecting their findings with previous knowledge (Goodwin and Goodwin, 1984a; Hammersley, 1985; Marshall, 1985; Silverman, 1989; Dingwall, 1992; Eisenhart and Howe, 1992; Silverman, 1993). The importance of cumulative research is illustrated by Dingwall and Murray's study of children in English A&E departments (Dingwall and Murray, 1983[53]). Although a number of participant observation studies had been carried out in A&E departments within the UK (e.g. Hughes, 1977; Jeffrey, 1979; Hughes, 1980), none of these had focused specifically upon children, in spite of their high use of A&E services. Dingwall and Murray took as their starting point Jeffery's analysis of the way in which patients are categorised in A&E departments. Jeffery had argued that doctors discriminate between two categories of patient: 'good' patients who allow doctors to practise the skills necessary for passing professional examinations, allow staff to practise their chosen speciality and test the general competence and maturity of staff, on the one hand, and 'bad' patients who were mostly those with trivial problems, drunks, overdoses and tramps. Jefferys observed that such bad patients received what could be described as punitive treatment including delay, inattention, verbal hostility and vigorous restraint. From their observations of children in A&E departments, Dingwall and Murray concluded that they routinely fulfil the criteria which Jefferys associated with being designated a bad patient. They were largely responsible for their own injuries, which often did not incapacitate them, were often able to extract secondary gains from their injuries and were often uncooperative.

However, contrary to what one would expect from Jeffery's conclusions, they received none of the punitive responses observed in relation to adult bad patients. As such Dingwall and Murray had identified a negative case, which was not accounted for by Jeffery's analysis. However, they went further than this. Drawing upon McHugh's proposal that commonsense conceptions of deviance should be understood as the outcome of judgements of the conventionality (i.e. avoidability of rule breaking) and theoreticity (i.e. understanding and intentionality of rule breaking) (McHugh, 1970), they proposed a fourfold typology of patient behaviour in medical settings, which serves to indicate those patient behaviours that will and will not attract punitive sanctions from health professionals. Using this typology, Dingwall and Murray were able to develop a coherent and dense analysis, which not only incorporated their own data and that of the other studies of A&E departments cited above, but also data drawn from studies of other medical settings. As they argued, they have, by building on earlier work without imposing its conclusions upon their analysis, shown how a consideration of the categorisation of children in A&E departments can modify, extend and elaborate earlier analyses.

5.3.7 Fair dealing

One of the commitments, which qualitative researchers working from either relativist or subtle realist assumptions share, is that any phenomenon may be understood from a number of different standpoints. This commitment to multiple perspectives (or in the case of relativists multiple realities) has serious implications for the truth claims of any research. In particular, the researcher must be wary of presenting the perspective of one group as if it this defined the objective truth about the phenomenon, while paying scant attention to other perspectives. Dingwall (1992) has called for fair dealing in research, which deals even-handedly with all those studied.

The issue of partisanship in research has been widely discussed in the literature. As Silverman (1993) noted, the discussion has tended to be somewhat polarised. On the one hand, there are those, such as Becker (1967a) who have argued that research is always morally and politically partisan and that therefore: "The question is not whether we should take sides, since we inevitably will, but rather whose side are we on?" (Becker, 1967a:239).

[53] See appendix 1 for details of this study.

Becker argued that research is inevitably carried out from a particular perspective, though this is not always obvious to either the researcher or the reader. It is particularly where the perspective of subordinates within an organisation is privileged, that the accusation of bias is most likely to be levelled. However, as Becker pointed out, privileging the definition of the situation that is put forward by superordinates is just as partisan, though it is much less likely to be called into question.

The commitment to partisanship, for example, in the work of Marxists such as Waitzkin (1979, 1990) (see above) and feminists, has been reflected in a commitment to advocacy on behalf of one or other oppressed group. On the other hand, there are those who have rejected such partisanship and have called for establishing facts through "the judicious testing of competing hypotheses and theories" (Silverman, 1993:172).

As Silverman pointed out, the latter position is untenable insofar as it fails to do justice to the way in which all research is inevitably shaped by values. However, the former is equally flawed, as it assumes that the recognition of the impossibility of value-free science forces us to take sides. Dingwall's call for fair dealing in research suggested an alternative position (Dingwall, 1992). In pursuit of error elimination the researcher should seek to incorporate interpretations from people at different status levels within a setting.

Becker (1967a) considered this possibility of including interpretations from people at different levels within an organisation or social setting. He dismissed it as posing the problem of **infinite regress**. He argued that: "Everyone has someone standing above him who prevents him from doing things just as he likes ... there is no end to it and we can never have a 'balanced picture' until we have studied all of society simultaneously" (247).

No doubt there is some truth in this. However, in choosing the focus of our research, we have chosen to bound the study in a particular way. It is surely reasonable to suggest that, within those boundaries, we should be as inclusive as possible.

This call for soliciting and incorporating different perspectives, drawn from different groups and interests, within a setting is echoed by a number of writers (Marshall, 1985; Silverman, 1985; Sandelowski, 1986; Guba and Lincoln, 1989). However, concerns about the particular groups that are seen as vulnerable to exclusion varies. Marshall (1985), Sandelowski (1986) and Guba and Lincoln (1989) are particularly concerned to avoid elite bias, the risk that the perspectives of the powerful might be over-represented, at the expense of the less powerful. This concern for the 'underdog' (Silverman, 1981), has a long history in qualitative research and much of the research carried out by the Chicago School in the 1930s focussed on underdog occupations (Fielding, 1993; Silverman, 1993).

More recently, however, a growing concern has emerged that this concern to do justice to the perspective of the underdog runs the risk of failing to represent the interactive character of social life because it does not deal even-handedly with the powerful and the privileged (Dingwall, 1992). Voysey (1975)[54] described this **underdog perspective** as a commitment to giving relatively powerless groups (in her case the parents of disabled children) "a chance to make themselves heard" (61). However, as she pointed out, there is little to differentiate such a position from that of the journalist. The danger is that a commitment to political or ideological correctness may lead the researcher to present the powerful and the privileged as cardboard cut-out characters, who are "either misguided or wilfully putting their own interests first" (Strong, 1979b:61). Strong (1988) described this as being more concerned with being 'right on' than with being right. Dingwall has acknowledged the seductiveness of the role of champion of the underdog, which pervades much qualitative research, not least in the health field where doctors and senior managers are easily cast in the role of oppressive villains. He commented:

> Such a role undoubtedly furnishes an element of romance, radical chic even, to liven the humdrum routine of academic inquiry. It is, however, inimical to the serious practice of ethnography, whose claims to be distinguished from polemic or investigative journalism must rest on its ability to comprehend the perspectives of top dogs, bottom dogs and, indeed lap dogs, and their respective contributions to the observable character of some organised social action. (Dingwall, 1980:874.)

Elsewhere, Dingwall has argued that: "Our science will never progress if we simply assume that all those white middle-class heterosexuals leading orderly lives represent some sinister force opposed to our underdog heroes or heroines and never acknowledge that they too are human beings making their way in an uncertain world" (Dingwall, 1992:172).

[54] See appendix 1 for details of this study.

Silverman (1993) has similarly criticised the self-righteousness of social researchers and has suggested that an insistence upon prioritising the pursuit of the emancipation of the underdog over the even-handed presentation of interpretations drawn from people at different status levels, can be seen as serving social scientists' own ends. He quoted Strong's scathing critique of those social scientists who seek to unmask the oppressive practices of others through their research: "In writing in this fashion sociologists both formulate themselves as members of some insightful and incorruptible elite and, at the same time, gain considerable pleasure by the exposure and thus potential overthrow of those whom they dislike" (Strong, 1979b:201).

Dingwall (1992) concluded that an important criterion in terms of which a qualitative research report should be evaluated is the extent to which it displays a commitment to fair dealing. For Dingwall, this marked the difference between social science and "muck-raking journalism" (172). He suggested that one of the questions we ought to ask in evaluating such a research report is: "Does it convey as much understanding of its villains as its heroes? Are the privileged treated as having something serious to say or simply dismissed as evil, corrupt or greedy without further inquiry" (172).

5.3.8 Summary

We have considered a number of ways in which the validity of qualitative research may be judged. We have considered the claims made for the use of respondent validation and triangulation and have concluded that, while both these methods may offer useful additional data for analysis, neither can be treated as a test of validity. We suggested that clear exposition of the processes of data collection and analysis, in which the data are related to the circumstances of their production, is essential to the evaluation of findings from qualitative research. We have argued that the risk of error will be reduced where the researcher pays systematic attention to the analysis of negative cases and to achieving fair dealing in the analysis and reporting of data.

5.4 Assessing the relevance of qualitative research

The second criterion proposed by Hammersley (1990)[55] and discussed above, was that of relevance. Hammersley argued that qualitative research must be relevant "however remotely" (107) to some public concern. Clearly, in HTA, the relevance needs to be somewhat less than remote. Nevertheless, as Hammersley argued, we must be wary of defining such relevance in terms which are too narrow. In particular, research needs to be directed not only at the issues that are identified as of immediate concern by practitioners, but also to those which there are reasonable grounds for believing are likely to be of relevance. Silverman (1993) argued that it is often unhelpful for researchers to take unquestioningly as their starting point a problem that has been identified by practitioners or managers. Not only may such problems reflect vested interests but:

> If field research has anything to offer, its theoretical imperatives drive it in a direction which can offer participants new perspectives on their problems. Paradoxically, by refusing to begin from a common conception about what is 'wrong' in a setting, we may be most able to contribute to the identification both of what is going on and, thereby, how it may be modified in pursuit of desired ends. (184–5)

Similarly, Dingwall (1992) pointed out that we cannot always take it for granted that the problem which practitioners identify is in fact the real problem which needs to be addressed. Part of the contribution which the cumulative body of social science can make is to recgonise the problems, which practitioners present, as potential members of a class of problems, for which particular research strategies or ways of formulating the problem have, in the past, proved to be useful.

Hammersley (1990) argued that relevant research should not only investigate issues that are significant, but should also make an original contribution to existing knowledge. The question to be asked here is whether research adds anything to what we already know about a topic. This is not to say, as Hammersley pointed out, that there is no place for confirmatory studies. One of the contributions which research can make is to test out what we already think we know and to put it beyond reasonable doubt (or otherwise). Research may also make a significant contribution by "plugging the gaps in our current knowledge" (115) and, just as importantly, calling into question what we have previously taken to be self-evident or checking that what was once established as true has remained so. Central to all of these functions is the idea that, to be relevant, research must in some way contribute to the accumulation of knowledge.[56]

[55] See section 5.2.3; [56] See sections 3.2.1.2 and 5.3.6;

Qualitative research has sometimes emphasised novelty at the expense of systematically building upon existing knowledge.

One of the key issues, which needs to be addressed in relation to the relevance of qualitative research, concerns the extent to which its findings can be generalised beyond the particular setting in which they were generated (Patton, 1980). While in certain circumstances, such as the evaluation of a particular programme, the value of a particular piece of research is not dependent upon the possibility of generalising beyond the setting in which the data were collected, this is not usually the case. Given that much qualitative research is based on a single or, at best, a small number of cases, this raises serious questions about the representativeness of findings (Bryman, 1988; Firestone, 1993; Silverman, 1993).[57] Some authors (Duffy, 1985; Marshall, 1985) have suggested that the uniqueness of every research setting, and qualitative methods (theoretical sampling, emergent design, large volume of data, etc.) are such that establishing the generalisability of findings beyond the settings studied is highly problematic. Others, however, while recognising the particular difficulties relating to generalisability in qualitative research, argue that it can and should be aspired to (Firestone, 1993). Hammersley (1980) identified two bases on which qualitative researchers may claim generalisability for their findings. These are empirical generalisation, on the one hand, and theoretical inference, on the other.[58]

As Hammersley (1992f) pointed out, probability sampling offers the most convincing basis for empirical generalisation and should be applied wherever possible within both qualitative and quantitative research (see also LeCompte and Preissle, 1993b).[59] However, such techniques are often inappropriate in qualitative research because the ratio of the settings which can be studied to the total number of settings to which one wishes to generalise is usually too low (see also LeCompte and Goetz, 1982; LeCompte and Preissle, 1993a). However, Hammersley argued, this does not mean that empirical generalisation is impossible within qualitative research. He cautioned against confusing probability sampling with the goal of making claims about typicality. Indeed, probability sampling methods are often not used in their pure form even in quantitative research, where stratified sampling techniques are used which rely on background information about the most significant ways in which the population of interests is heterogeneous. In qualitative research, establishing the empirical generalisability of findings depends upon the researcher demonstrating the likelihood that the setting or settings studied are representative of the population to which (s)he wishes to generalise (Beck, 1993). Hammersley recommended three sources of information, which may be used to do this. First, the researcher may be able to make use of published statistics to establish the similarities between the setting studied and the aggregate to which (s)he wishes to generalise. Second, particularly where published statistics are not available to support or refute such claims to typicality, the researchers may be able to combine case studies with survey work in the same investigation in such a way as to collect data of relevance. Third, it may be possible to coordinate a series of ethnographic studies of the phenomenon of interest (Bryman, 1988), in a range of different studies, which might be selected on the basis of stratified sampling. For example, if one were studying interactions between GPs and other professional groups in general practice, one might identify a number of dimensions along which such practices vary (urban and rural; fundholding and non-fundholding; single handed and group practices; purpose built and converted premises etc.) and use these as the basis of a stratified sample of cases to be studied.

A number of authors (e.g. Lincoln and Guba, 1985; Henwood and Pidgeon, 1992) have argued that generalisability, as conceived of in quantitative science, is an impossible goal for qualitative research. Guba (1981) suggested that: "Generalizations of the rationalistic variety are not possible because phenomena are intimately tied to the times and contexts in which they are found" (81).

They proposed the more modest goal of **applicability** (Guba, 1981) **transferability** (Lincoln and Guba, 1985) or **case-to-case transfer** (Firestone, 1993). Transferability represents the extent to which the findings of a particular study may be applied to similar contexts. The crucial issue here is how such similar contexts are to be defined. Unlike Hammersley (1992f), who insisted that it is the researchers' responsibility to identify the population to which they are claiming that their findings may be generalised, Lincoln and Guba (1985), Walker (1980), Patton, 1980 and Kennedy (1979) have proposed that the responsibility for deciding about the transferability of research findings lies

[57] See section 4.1 for a fuller discussion of generalisability in qualitative research;
[58] See section 4.1 for a detailed discussion of these approaches to sampling; [59] See section 4.1.1.1.

with the reader of the research, rather than the researchers themselves: "It is the reader who has to ask, what is there in this study that I can apply to my own situation and what clearly does not apply" (Patton, 1980:34).

The original researcher bears the responsibility for providing sufficient descriptive detail to allow the reader to make an informed judgement about whether it is likely that the original setting and the setting of interest are sufficiently similar to give reasonable confidence that the findings in one setting will hold in the other (Kennedy, 1979; Merriam, 1988). They call for "thick description" (Walker, 1980) of the sending context to allow readers to consider potential comparability (see also LeCompte and Preissle, 1993c). Guba (1981) argued:

> If the thick descriptions demonstrate an essential similarity between the two contexts, then it is reasonable to suppose that tentative findings of Context A are also likely to hold in Context B (although, to be safe, an empirical test of that assumption should be made). ... The naturalist does not attempt to form generalizations that will hold at all times and in all places, but to form working hypotheses that may be transferred from one context to another depending upon the degree of 'fit' between the contexts.

Similarly, Jensen (1989) called for explicit detailing of methods, definitions and theoretical frameworks to allow translation into other settings. LeCompte and Preissle (1993c) proposed that the researcher must demonstrate the typicality of the phenomenon of interest, by identifying all the dimensions of the phenomenon which are considered salient.

The second basis of generalisation in qualitative research, discussed by Hammersley (1992f), is theoretical inference. Firestone (1993) referred to this as **analytic generalisation**. Silverman (1993) argued that it is important to recognise that, in qualitative research, generalisation does not depend entirely upon a statistical logic. Where, as often is the case with qualitative research, data is drawn from a single setting, inferences from that settings to other settings cannot be statistical and must depend upon the adequacy of the theory on which they are based (Silverman, 1989). Bryman (1988) made the same point when he argued that: "The issue [of generalisability] should be couched in terms of the generalisability of cases to theoretical propositions rather than to populations or universes" (90).

Bryman took the classic study of the care of the dying in hospital, which was carried out by Glaser and Strauss in the 1960s, to illustrate his point (Glaser and Strauss, 1965b). In this study, Glaser and Strauss developed a substantive theory, on the basis of their study of a hospital, which suggested that the greater the social loss of a dying patient, the better the care received and the more likely nurses are to produce rationales to explain the death. They translated this into a formal hypothesis: "The higher the social value of a person the less delay he experiences in receiving services from experts" (42). At this point, the issue was not whether the experiences of dying patients in the hospital studied were typical of such patients in other hospitals, but whether the experiences of such patients were representative of those of the class of people who receive services from experts. By moving the generalisation from the concrete (or substantive) level, it became possible to test its validity by examining it in other settings, which belonged to the same theoretical class (people who receive services from experts). This pointed the way to studies of other settings, such as a study of the significance of the social value of the patients who attend their general practice surgeries with chronic disease for the speed with which they will receive treatment. Attempts to establish such theoretical generalisations on the basis of further empirical work will make use of theoretical sampling,[60] analytic induction,[61] and the pursuit of negative cases,[62] as discussed above. Mitchell (1983) argued that this is the only basis on which generalisation from single cases is appropriate. To attempt empirical generalisation, on the basis of case studies, is, he argued, to confuse the logic of enumerative induction, which makes use of statistical inference, with that of analytic induction.

5.4.1 Summary

We have argued, with Hammersley (1992) that relevance is one of the two criteria by which research should be assessed. The relevance of a particular piece of research, qualitative or otherwise, will always be a matter of judgement and we have cautioned against too facile a definition of relevance. The generalisability of findings from qualitative research poses significant problems for its relevance. We have suggested three, in our view, compatible ways in which qualitative research may support its claim to generalisability beyond the setting in which it was produced. First, qualitative research, particularly interview studies, may be able to make greater use of statistical sampling

[60] See section 4.1.2.4; [61] See section 4.6; [62] See section 5.3.6.

techniques than they generally do. Secondly, claims to case-to-case transfer will be strengthened where the researcher provides the reader with a full description of the sending case. Opinions vary about whether the researcher or the reader should take responsibility for demonstrating the similarities between the sending and receiving context, but in either case a full description of the original setting will make an important contribution to demonstrating the potential generalisability of findings. Finally, the logic of analytic induction, which seeks to systematically establish the potential for and limits to the generalisability of theoretical claims arising from and within particular settings offers an alternative approach to generalisability.

5.5 Conclusion

It is clear from the above discussion that there is considerable disagreement among qualitative researchers about whether any criteria may be applied to their products, about which criteria are appropriate and about how research may be assessed in relation to such criteria. Following Hammersley (1992e), we have proposed that the most appropriate criteria for evaluating qualitative research are validity and relevance. We have argued that, while relevance is always a matter or judgement, and validity can never be established beyond all doubt, there are a number of ways in which error may be limited and the quality of research should be judged in relation to the extent to which the researchers have succeeded in such error limitation.

Implications for commissioning and practice

- The same criteria are appropriate for both qualitative and quantitative research in HTA.
- The criteria against which both qualitative and quantitative research in HTA should be evaluated are validity (the extent to which the likelihood of error has been limited) and relevance. If research is to be used to inform the development and application of health technologies and to assess their impact, we must have some confidence that the findings are true. If such research is to be useful its relevance must be clear.
- Given the differences between qualitative and quantitative research, in terms of the problems which each addresses and the research practices employed, the appropriate means for judging validity also differ.
- It is impossible and inappropriate to devise rigid checklists for assessing the validity of qualitative research findings. The validity of any research is always a matter of judgement.
- While respondent validation exercises may provide additional data, they cannot be treated as tests of the validity of research findings.
- The use of multiple methods in research in HTA can increase the comprehensiveness of a study and may also stimulate reflexive analysis of data from different sources. However, there are then very significant problems with adopting triangulation as a test of the validity of research findings.
- Judgements about the validity of research depend upon being able to form a judgement of the research process which led to the findings presented and hence evaluate the evidence upon which claims are based. It is therefore incumbent upon researchers to create and present in their research reports a record of the research, which includes an adequate account of the process by which findings were produced. Given the non-standardised nature of qualitative research such records are likely to be more elaborate than in reports of quantitative research.
- Researchers should produce and present clear definitions of the concepts and categories which are developed in the course of the research, so that their meaning and application is unequivocal.
- Researchers should demonstrate that the conclusions they have drawn are justified in relation to the data collected.
- Confidence in the accuracy of the data collected is increased where the researchers have used mechanical means (such as audio and/or video recording) for recording, and standardised procedures for transcribing, their data. However, the use of such devices and techniques is not always appropriate or practical.
- Researchers should display enough data to allow the reader to assess whether the interpretations are adequately supported by the data. Such data should be separated out from the researchers' analysis of them. The use of low inference descriptors such as verbatim accounts and precise descriptions in research reports allows the reader to form a judgement about the validity of the analyst's interpretation.
- Confidence in the validity of research findings will be increased where researchers have spent an extended period in the setting under study since this allows the researcher to search thoroughly for disconfirming data and to identify ways in which the researcher's presence has contributed to the data obtained.
- The trustworthiness of researchers' analyses of their data is enhanced where researchers can demonstrate that they have considered alternative plausible explanations for their data.

- Confidence in the validity of researchers' findings will be increased where there is evidence of sensitivity to the ways in which the researcher's presence in the research setting has contributed to the data collected and their own *a priori* assumptions have shaped the data analysis.
- The likelihood of error is reduced where researchers demonstrate that they have engaged in a conscientious search for and presentation of cases that are inconsistent with their emerging analyses. This can be achieved through the systematic use of theoretical sampling and searches for negative or disconfirming cases.
- The validity of research findings is enhanced where the researchers increase our understanding of all members in a setting and do not present one-sided accounts from the perspectives of either elites or underdogs.
- While qualitative research in HTA must be relevant, it is important to avoid defining such relevance too narrowly. One of the strengths of qualitative method is that it can offer participants new perspectives on their problems. It is therefore often unhelpful for qualitative researchers to adopt unquestioningly the perspectives of practitioners, managers or policy makers as their starting point.
- The relevance of qualitative research is increased where design and analysis take into account and add to what we already know about a field.
- The relevance of research in HTA is related to its potential generalisability to groups or settings beyond those studied. Given that most qualitative research is based on a single case or, at best, a small number of cases, the generalisabillity of qualitative research is a particular concern.
- While probability sampling is rarely practical, this does not mean that concerns with generalisability should be abandoned. Rather researchers should seek to enhance the empirical and theoretical generalisability of their findings.

Chapter 6

Using qualitative research methodologies to study medical information systems

This case study examines the implications of the material reviewed in chapters 3, 4 and 5 for the assessment of one specific area of health technology – that of the use of computer systems in healthcare settings. This case study illustrates the potential of qualitative research in HTA. First we discuss some of the limitations of quantitative research in this field and the complementarity of qualitative and quantitative approaches (see chapter 3). Qualitative research on information systems illustrates the difference, discussed in chapter 4, between interactionist research that seeks to explore how participants make sense of new information systems and ethnomethodological research, which focuses upon what participants do rather than upon such meanings. These differences, and the strengths and weaknesses of each approach are examined through the detailed discussion of two studies, one rooted in each of these traditions. In presenting these two studies, we show how some of the quality markers, discussed in chapter 5, have been applied, in practice, in qualitative research in HTA.

The studies reviewed here have been chosen to illustrate many of the principal issues, which are dealt with in the earlier part of the report. The links between these issues and the rest of the report are made explicit in section 6.6.

Although computer information systems have the potential to improve the efficiency and quality of the delivery of healthcare services, many systems do not achieve these objectives. Often, despite being technologically sound, they are reported to be inadequate, ineffective and/or underused. Increasingly, it is recognised that these problems arise because developers emphasise the technological and economic aspects of systems, while neglecting social considerations, such as the perspectives and work practices of the people who use them. Thus, for example, surveys of hospitals in the USA reveal that up to 50% of information systems fail due, at least in part, to staff resistance and interference (Dowling, 1980; Lyytinen, 1987; Lyytinen and Hirschheim, 1987).

The success of clinical information systems, therefore, depends not only on their being technologically sound, but also on complex social processes, which affect, and are affected by, their introduction. Recognising this, a number of researchers have turned their attention to the social aspects of the implementation of computer systems in healthcare settings (Anderson *et al,* 1994a; Greatbatch *et al,* 1995b). As Anderson and co-workers noted, these researchers have shown that the social sciences provide: "theoretical frameworks and analytical approaches that can help understand how the introduction of computer systems in healthcare settings affects the quality of the work environment, tasks and skills of health professionals, social interactions among professionals in the organisation, and the effective delivery of medical care" (Anderson *et al,* 1994b:ix).

Some social scientific studies of the implementation of technologies in healthcare settings use quantitative methodologies to assess their impacts. Others, however, use qualitative methodologies which may be less familiar to professionals involved in the development, deployment and use of medical information systems. The purpose of this case study is to show how qualitative approaches contribute to research on the organisational impacts of healthcare technology. We begin by introducing several critical issues, which can be most effectively investigated through the use of qualitative methods. We then identify two forms of qualitative research and illustrate each of these with exemplary studies of the use of particular systems. We conclude by discussing how qualitative research can contribute to the design and deployment of healthcare technology.

6.1 Approaches to researching medical information systems

6.1.1 Quantitative research on information systems

Most information systems research involves the formulation and testing of hypotheses "through controlled experiment and/or statistical analysis" (Kaplan and Duchon, 1988:572). Usually such

research aims to measure the quantitative outcomes of the implementation of computer systems. These outcomes include the effectiveness of systems, their impact on the cost of services or treatments, and user satisfaction.

Quantitative studies have provided important insights into the implementation of computer systems, especially in relation to technical and economic aspects of system performance. Recently, however, researchers in several disciplines have suggested that the dominance of the positivist perspective within information systems research has led to the neglect of social and cultural processes surrounding the development and implementation of computer systems (e.g. Forsythe and Buchanan, 1992; Jirotoka and Goguen, 1994). According to this view, although quantitative methods are appropriate for assessing factors that can be conceptualised as discrete entities, such as rates of error and usage, they exclude from consideration a range of "phenomena that do not lend themselves to quantification" (Forsythe and Buchanan, 1992:10). In particular, they do not provide access to the dynamics and complexities of the social processes, meanings, contexts and practices surrounding the use of information systems over time, in real-world settings.[1] As Forsythe and Buchanan stated, qualitative research is particularly well-equipped to address these features:

> For the evaluation of non-technical aspects of system functionality and acceptability, the methods of qualitative social science are more suitable (than quantitative and experimental approaches). Such unobtrusive methods as participant observation and interviewing can provide systematic data on patterns of thought and behaviour in natural workplace settings. (Forsythe and Buchanan, 1992:9.)

Although many positivists recognise the importance of qualitative methods, they often see them as a means of generating hypotheses for quantitative research and/or of contextualising the findings of such research.[2] Below we suggest that qualitative methods should be assigned a central role in information systems research.

6.1.2 Qualitative research on information systems

Recently, qualitative methods have been used to study the use of information systems in numerous organisations, including hospitals and medical centres, urban transport control rooms, air traffic control centres, and financial dealing rooms. Such studies have examined the social processes surrounding the development, implementation and use of information systems over time, in natural settings. Although the studies employed a variety of approaches, they can be separated into two broad groups: interactionist and ethnomethodological. Interactionist studies share the goal of understanding how participants construe and conceptualise computer information systems and the activities in which they are implicated.[3] By investigating differences and similarities in the perceptions of various individuals and groups, these studies have described how systems affect and/or are affected by the inner workings of the organisations in which they are developed, deployed and utilised. This is usually achieved through in-depth case studies of particular systems, using data gathered via field observation, interviews, questionnaires and/or examination of organisational documents. The methods used in these studies are discussed by Shneiderman and Carroll (1988) and Forsythe and Buchanan (1992).[4]

In contrast, ethnomethodological studies are not primarily concerned with the participants' points of view.[5] Instead they focus on the **practices** that participants use in developing, installing and using computer systems. Thus, whereas interactionists concentrate on how people **see things** (i.e. on the expectations and evaluations that people communicate to one another), ethnomethodologists focus on how they **do things** (Maynard, 1989); they describe the routine, often tacit practices that participants use in accomplishing workplace activities. Again, these studies involve in-depth analyses of particular systems and work settings. While some rely solely on the same sources of data as interactionists (e.g. Blomberg, 1995; Hughes *et al,* 1994; Randall *et al,* 1994), there is an increasing emphasis on the use of audio-visual recordings of work and interaction in real world settings. These data, it is argued, offer a number of advantages over more traditional modes of inquiry such as field work, interviews and questionnaires (e.g. Heritage, 1984; Heath and Luff, 1993). They provide access to the intricacy of workplace activities by enabling researchers to repeatedly examine particular events. Unlike field notes and responses to

[1] See section 3.2.2.4 for a discussion of the role of qualitative research in studying the processes involved in introducing technologies; [2] See sections 3.2.1.2 and 3.2.2.5 for a discussion of the role of qualitative research in generating hypotheses; [3] See section 3.2.2.1; [4] See also Cash and Lawrence (1989), Mumford *et al* (1985), Nissen *et al* (1991), Walsham (1993); [5] See sections 2.4.4. and 4.5.

questions, they also provide raw data, which can be re-examined in the light of new findings and in terms of different research questions and interests. Moreover, by enabling researchers to evaluate the strength of each other's analyses by reference to the raw data, they provide an important constraint on the influence of personal preconceptions and analytic bias.

This is not to say that researchers who specialise in video-based analysis of workplace activities have abandoned more traditional forms of qualitative data. Often their analyses of video data are informed by information gathered through observations, interviews and questionnaires. This is because the patterns of interaction and the relationship between activities in complex organisational settings often cannot be discerned or understood solely from analysis of video recordings. The methods used in these studies are discussed in Greatbatch and co-workers (1995b), Heath and Luff (1993) and Luff and co-workers (1994).

To date there have been relatively few qualitative studies of the social processes surrounding the use of healthcare information systems. To illustrate the potential of such research, we discuss and contrast an interactionist and an ethnomethodological study of the use of computers in medical settings. These studies show that qualitative research provides insights into aspects of organisational conduct which are largely inaccessible to quantitative research.

6.2 Interactionist studies: investigating participants' perspectives

A number of interactionist studies have been conducted in North America. These studies investigated the impacts of several systems in both primary and secondary health care. The systems include technologies designed for use throughout hospitals by a variety of staff (Fischer *et al,* 1980; Lundsgaarde *et al,* 1981), in pharmacies (Aydin, 1989; Aydin and Ischar, 1989), in clinical laboratories (Kaplan, 1986; 1987; Kaplan and Duchon, 1988, 1989), and in neurology departments (Nyce and Graves, 1990). The insights provided by interactionist studies of computer use can be summarised as follows:

- Understanding how medical professionals conceptualise, perceive and understand computer information systems.[6] To understand and explain people's behaviour with respect to computer systems, it is necessary to investigate how they understand and interpret such systems in natural settings. These subjective phenomena cannot be studied adequately through purely quantitative approaches because, amongst other things, they vary over time and often display sensitivity to local circumstances and contingencies in ways that can only be grasped through observation, informal interviews and related qualitative data. Consequently, qualitative research can make an important contribution to explaining a system's successes and failures.
- Understanding how computer systems affect, and are affected by, the organisational contexts in which they are developed, installed, and used.[7] Qualitative methods enable researchers to improve our understanding of the contexts into which systems are introduced. Contextual factors include the professional values and philosophies of systems designers and users, patterns of communication between occupational groups and changes in the status of medical personnel due to the introduction of new systems.
- Investigating causal processes.[8] By providing insights into the inner workings of organisations and occupational groups, qualitative research can explain causal processes identified by quantitative studies. Thus, as Kaplan and Maxwell observed, although quantitative studies:

> … can demonstrate **that** causal relationships exist, they are less useful in showing **how** causal processes work. Qualitative methods often allow the researcher to get inside the black box of experimental and survey designs and to discover the actual processes involved. Qualitative research is particularly good for developing explanations of the actual events and processes that lead to specific outcomes. In this way, qualitative methods can yield theories and explanations of how and why processes and outcomes occur. (Kaplan and Maxwell, 1994:48.)

To illustrate the interactionist approach, we consider a study of the use of a computer information system in clinical laboratories within an academic medical centre in the USA (Kaplan, 1986, 1987; Kaplan and Duchon, 1988, 1989).

6.2.1 An interactionist study of the use of a clinical laboratory computer information system

Kaplan and her colleagues investigated the impacts of the installation of a computer

[6] See section 3.2.2.1; [7] See section 3.2.2.3; [8] See section 3.2.2.4.

information system by all nine clinical laboratories within a 650-bed university medical centre. Their study is specially useful in that it involved both quantitative and qualitative methods and it clearly identified the strengths of the latter. In particular, it showed how qualitative analyses centred on users' perspectives can reveal information that is crucial for understanding patterns of system use in medical settings.

6.2.1.1 Research setting and the system

The clinical laboratories conduct tests ordered by physicians to diagnose or track the course of illness and disease. The tests include "blood sugar measurements (and) assessments of bacterial sensitivity to antibiotics" (Kaplan and Duchon, 1988:575). They are carried out by laboratory technologists who are also responsible for reporting the results.

The computer information system, which was installed in April 1985, replaced a manual data management system. According to Kaplan and Duchon:

> The principal function of computers in clinical laboratories involves data management. Computers can relieve the clerical burden of data acquisition and transcription while adding new data entry and computer-related tasks. In addition, they improve legibility, organisation, and accuracy of laboratory results reports; increase productivity and efficiency; reduce transcription error; and change laboratory organisation and turnaround time for test results. Thus, such computer information systems affect both the process of work as well as the service product of the laboratory. (Kaplan and Duchon, 1988:575.)

6.2.1.2 Methods

The research team was interested in the interrelationships between the computer system and the work of the technologists. However, the quantitative and qualitative researchers approached this topic in different ways. The quantitative researchers assessed the impact of the system on laboratory work through statistical analysis of quantitative survey data. In so doing, their main interest was in testing pre-existing theory about job characteristics and job satisfaction in a new setting. In contrast, the qualitative researcher determined the laboratory technologists' reactions to the system by examining qualitative data gathered from several sources (see below). Moreover, she developed and reformulated her hypotheses and theories inductively[9] from the data as her research proceeded.

Qualitative data were gathered from "open-ended interviewing, observation, participant observation, and analysis of responses to open-ended items on a survey questionnaire" (Kaplan and Duchon, 1988: 577). Kaplan and her associates began by interviewing several people before the installation of the computer information system, including the departmental chair, the hospital administrator, and the laboratory directors. These interviews were designed:

- to determine what interviewees expected the potential effects of the computer system to be on patient care, laboratory operations, and hospital operations
- to inquire about possible measures and focus of the study
- to generate questionnaire items for a survey of the laboratory technologists (Kaplan and Duchon, 1988:577).

After, the installation of the system the researchers spent time in the laboratories observing and questioning the laboratory staff while they worked. Kaplan also attended – and occasionally participated in – departmental meetings where directors, supervisors and others discussed system problems.

The quantitative researchers used these qualitative data as a resource to develop quantitative measures for their survey questionnaire. They were not able to conduct a pre-installation survey because access to the research site was not secured until just before the introduction of the system. The first questionnaire was administered to the laboratory technologists 7 months after the system was installed. A modified version of the questionnaire was administered "approximately one year later, when the initial changes caused by the computer system became part of normal procedure" (Kaplan and Duchon, 1988:577).

The questionnaire included four open-ended questions concerning the impact of the system and how system use could be improved. These questions provided technologists with an opportunity to raise issues that were not covered by the questionnaire and/or "for which measures were difficult to develop" (Kaplan and Duchon, 1988:577). The responses to these questions were subjected to qualitative analysis by Kaplan.

6.2.1.3 Initial qualitative analysis

Kaplan's objective "was to identify and account for both similarities and differences among laboratory

[9] See section 3.2.1.2.

technologists and among laboratories in their responses to the computer information system" (Kaplan and Duchon, 1988:576). To achieve this objective, she used the constant comparative method (Glaser and Strauss, 1967)[10] to analyse her interview data and the technologists' responses to the open-ended questions on the questionnaire. Using this method, she aimed to derive categories that reflected the perceptions of the laboratory directors and the technologists. Kaplan and Duchon described the constant comparative method as follows:

> [It involves] an interpretative approach that uses data to both pose and resolve research questions. Researchers develop categories and meanings from the data through an iterative process that starts by developing an initial understanding of the perspectives of those being studied. That understanding is then tested and modified through cycles of additional data collection and analysis until coherent interpretation is reached. Thus although qualitative methods provide less explanation of variance in statistical terms than quantitative methods, they can yield data from which process theories and richer explanations of how and why processes and outcomes occur can be developed'. (Kaplan and Duchon, 1988:573.)

Kaplan began by analysing the technologists' responses to the open-ended questions on the questionnaire. She found that:

> Three themes predominated in the answers: (1) changes in technologists' work load, (2) improvements in results reporting, and (3) the need for physicians and nurses to use computer terminals rather than telephones for results inquiry. Technologists expressed a general sense that their clerical duties and paperwork had increased and productivity had suffered. However, they credited the computer system with making test results available more quickly. They said that results reports were also more complete, more accurate, easier to read, and provided a picture of 'the whole patient'. Even though phone calls interrupted lab work, they felt that doctors and nurses expected to get tests results by calling the labs, rather than by using the computer system. In addition, respondents sensed they were being blamed by others in the medical centre for problems caused by the computer system. (Kaplan and Duchon, 1988:578.)

Kaplan's analysis also indicated that individual laboratory technologists, as well as the laboratories as a whole, differed in their assessments of the impacts of the computer system. Some focused on increased workloads, whereas others stressed improvements in the reporting of test results.

6.2.1.4 Initial quantitative analysis

To measure the technologists' attitudes to the impacts of the system on their work, the quantitative researchers employed standardised quantitative measures of job characteristics. Their analysis of the quantitative data, which used a standard statistical software package, did not reveal any differences between the technologists' attitudes to the impacts of the system on their workloads, results reporting, and communications with the rest of the medical centre. In other words, in contrast to the qualitative study, the quantitative study detected no differences between the technologists' in terms of reactions to the computer system.

6.2.1.5 Subsequent qualitative analysis

Kaplan was convinced that her initial analysis of the interview data and open-ended questions was valid and that the quantitative study had failed to pick up important differences between individual technologists and the laboratories: her analysis was supported by information obtained from her fieldwork. Therefore, she sought to account for the differences and to explain why the quantitative study had not detected them.

Kaplan began by reanalysing the interviews with laboratory directors, in order to establish their expectations prior to the installation of the system. She thought that "perhaps different expectations among directors could have contributed to different responses within the laboratories" (Kaplan and Duchon, 1988:579). However, she found that "prior to implementation, directors generally agreed that there would be more work for laboratory technologists, but that nature of technologists' jobs would not be changed by the computer system" (Kaplan and Duchon, 1988:579). Thus the differences could not be explained by the contrasting expectations of directors about the system.

Further analysis of the qualitative data provided an alternative explanation. Specifically, Kaplan found that:

> different technologists had different views of their jobs, and these different views affected their attitudes toward the computer system. (Kaplan and Maxwell, 1994:49.)

> One group saw their jobs in terms of producing results reports, the other in terms of the laboratory bench work necessary to produce those results reports. The

[10] See section 4.6.1.4.

> group who saw its jobs in terms of bench work was oriented towards the process of producing lab results, whereas the group who viewed its work in terms of reporting results was oriented towards the outcomes of the lab work; the members of this group saw themselves as providing a service. (Kaplan and Duchon, 1988:580.)

Kaplan's analysis indicated that the group who saw their work in terms of the laboratory bench work focused on impacts concerning this part of the technologist's job. That is, they concentrated on the ways in which the system affected the process of **producing** laboratory results; for example, increasing/decreasing paperwork, the time needed for data entry, and numbers of telephone calls. Generally, this group viewed the system as increasing their workload and interfering with their work. In contrast, the group who defined their work in terms of the product, rather that the process of laboratory work, assessed the system in terms of how it affected result reporting. On the whole, this group offered favourable assessments of the computer. They saw it as improving the service they delivered to the rest of the medical centre, and thus as a boon rather than a burden. Thus the technologists' different perceptions of their jobs explained why for some technologists, the system enhanced their jobs, whereas for others it interfered with their jobs, even though they ostensibly had the same jobs and were using the same system (Kaplan and Maxwell, 1984:49).

These differing perceptions explained why the initial quantitative analysis did not discern differences among laboratory technologists responses to the system.

> although the two groups of lab technologists differed according to how they viewed their jobs, these kinds of differences in job orientation would not be measured by the [standardised] job characteristics measures [used by the quantitative researchers]. Job characteristic measures do not assess or measure such particularistic aspects of how respondents define their jobs. Consequently, there need be no correlation between job satisfaction and supervisory status, education, previous employment, laboratory, and computer system measures. (Kaplan, 1987)

In the light of these findings, subsequent analysis of the quantitative data incorporated two new variables to establish whether the technologists' responses differed according to whether they emphasised the process or the service aspects of their jobs. This analysis supported the results of the qualitative study.

6.2.2 Discussion

In sum, this study involved two dimensions: a quantitative study involving a survey questionnaire designed to assess the impact of the computer system on work in the laboratories; and a qualitative study involving field observations and open-ended questionnaires designed to establish what changes participants attributed to the computer system. The quantitative study initially found no differences among the responses of the users. In contrast, qualitative analysis of open-ended interviews revealed that the users differed markedly in their reactions to the introduction of the computer system. Some emphasised increased workloads, whereas others emphasised improvements in efficiency. These differences were subsequently found to stem from differences in the ways personnel viewed their jobs. The latter had eluded the statistical analysis because the standardised measures used by the quantitative researchers assumed a uniformity of job situations and perceptions, which did not exist in practice.

This study indicates the importance of qualitative research in investigating the installation and use of information systems and other technologies. To evaluate the impacts of computer systems, it is important to understand the perspectives of those involved with the systems and the social relationships and patterns of communication within they are enmeshed. Indeed, in the present case the researchers would have been seriously misled had they relied on quantitative procedures. For as Kaplan noted:

> it was not possible to design, in advance, a quantitative study that would have tested the right hypotheses, because appropriate hypotheses could not be known in advance. A qualitative approach enabled the researchers to see how individuals construed the information technology, their jobs, and the interaction between the laboratory computer information system and their jobs. Thus the researchers were able to generate productive hypotheses and theory. (Kaplan, 1987:50)

Thus, the study by Kaplan and co-workers illustrates the strengths of interactionist research and shows how quantitative and qualitative studies can work in concert. It also shows that the latter should not be viewed merely as a precursor to the development of standardised questionnaires. If we wish to identify and account for the impacts of information systems and to understand why systems are accepted or rejected, then in-depth qualitative analysis is essential.

6.3 Ethnomethodological studies: investigating participants' practices

Ethnomethodological studies of the use of new technologies have largely been undertaken in non-medical settings (Goodwin and Goodwin, 1997; Heath and Luff, 1992a, 1992b; Hughes *et al,* 1994; Suchman, 1987; Whalen, 1992; 1995; 1996). However, these studies suggest that ethnomethodological research could improve our understanding of medical information systems by showing how their use (or non-use) affects and is affected by the work practices of medical personnel. To illustrate this, we consider a study of the use of a computer information system in an inner-city general medical practice in England.

6.3.1 An ethnomethodological study of the use of a general practice computer information system

This study examined the ways in which doctors and patients interact with a computer system within the general practice consultation (Greatbatch *et al,* 1993; Greatbatch *et al,* 1995a; 1995b). Like Kaplan's research, it illustrates the importance of the social processes surrounding the use of new technologies in healthcare settings. In contrast to the Kaplan, however, the researchers are not primarily interested in ascertaining the perspectives of the participants. Instead, they focused on the practices that the participants used during the course of diagnostic and prognostic activities within the general practice consultation.[11]

6.3.1.1 Research setting and the system

The study focused on the installation and use of a computer information system in a inner city practice in North West England. The computer information system, known as Value-Added Medical Products (VAMP, after the company that produces it), was installed in 1990. VAMP was designed to support two interrelated activities within primary health care. First, it allows the doctor to enter and retrieve details of patients' complaints and treatments. Second, it provides a facility for issuing prescriptions for treatment to patients. The system is run on conventional desk-top computers consisting of a monitor and keyboard. Information is entered by using the keyboard and following a series of prompts which require the doctor to move progressively through a series of options. Prior to the installation of the system in 1990, the patients' medical records were stored on paper cards and prescriptions were written by hand.

6.3.1.2 Methods

The study by Greatbatch and co-workers (1993; 1995a; 1995b) was concerned with the impact of computerisation on professional practice and interpersonal communication within the general practice consultation. It involved the qualitative analysis of video recordings of medical consultations conducted before and after the introduction of VAMP into the research setting. Their analytical framework was derived from CA,[12] a form of ethnomethodological[13] research which uses inductive search procedures to locate patterns of verbal and/or non-verbal conduct in audio and audio-visual recordings of naturally occurring activities. The objective of CA is to describe the practices which people use and rely upon in producing their own behaviour and interpreting and dealing with the behaviour of others. The central resource out of which analysis emerges is the moment-to-moment understandings that participants unavoidably display as they interact with one another.

In locating and analysing recurring patterns of action and interaction, CA researchers repeatedly replay their audio or video recordings of naturally occurring interaction, carefully transcribing the events. The transcripts capture not only what is said, but also a variety of details of speech production and, in the case of video data, visual conduct such as gaze direction and gesture. These transcripts facilitate the fine-grained analysis of the recordings, enabling researchers to reveal and analyse tacit, taken-for-granted aspects of human conduct. Extracts from the transcripts are included in research reports as examples of the objects of inquiry.

The following is a simplified version of the notation used in CA studies:

Transcription notation[14]

[A left bracket indicates the point at which overlapping talk begins.

] A right bracket indicates the point at which overlapping talk ends.

= Equals signs indicate that different speakers' utterances are **latched**. They also link continuous talk by a single speaker that has been distributed across non-adjacent utterances because of another speaker"s overlapping utterance.

[11] See section 4.5; [12] See section 4.5 for a detailed discussion of conversation analysis; [13] See section 2.4.4 for a discussion of ethnomethodology; [14] For a more extensive glossary, see Atkinson and Heritage (1984).

(0.5) Numbers in parentheses indicate the length of silences in tenths of a second.
(.) A dot in parentheses indicates a gap of less than two-tenths of a second.
– A dash indicates a cut-off sound like a guttural stop.
Word Underlining indicates some form of stress via pitch and/or amplitude.
WORD Capital letters indicate utterances, or parts thereof, that are spoken much louder than the surrounding talk.
Wo::rd Colons indicate prolongation of the immediately preceding sound.
. , ? Periods, commas, and question marks are used to indicate falling, non-terminal, and rising intonation, respectively.
(Word) Parenthesised words indicate that the transcriber was not sure of what was said.
( ) Empty parentheses indicate that the transcriber could not hear what was said.
(( )) Double parentheses contain transcriber's comments and/or descriptions.
.hhh hs preceded by a period represent discernible inhalations.
hhhh hs without a preceding period represent discernible aspiration.
____ A continuous line above a speakers talk indicates that the speaker is gazing at the co-participant.
,,,, A series of commas above a speakers talk indicates that the speaker is turning away from a co-participant.
.... A series of dots above a speakers talk indicates that the speaker is turning to another.

The data used in the study by Greatbatch and co-workers (1993; 1995a; 1995b) consisted of video recordings of medical consultations conducted before and after the installation of VAMP, augmented by field observation and discussions with the doctors. Approximately 100 consultations were recorded before the installation of the computer system and 150 after its installation. The recordings were collected at regular intervals so that the researchers could detect changes in the use and effects of the computer system as the doctors became more familiar with its operation, constraints and potential. Seven GPs were represented in the recordings, four of whom were involved in each phase of recording.

Data analysis and results. Greatbatch and co-workers (1993; 1995a; 1995b) began by considering the relationship between the use of the computer system and the interaction between doctor and patient. They then compared consultations conducted before and after the introduction of the system. This revealed that the installation of the system had affected the doctors' ability to enter and retrieve information flexibly during the consultation.

Computer use in the consultation. While reviewing the recordings of consultations involving the use of a computer, Greatbatch and co-workers noted the extent to which the use of the system appeared to be interrelated with communication between doctor and patient. First, they noticed that the use of the system was frequently coordinated with the interaction between doctor and patient. This ranged from brief cessation of system use as a doctor glanced at a patient through to collaborative readings of screen-based text, some of which were initiated by the patients. Second, they found that the doctors also regularly coordinated their social interactional conduct with the use and operation of the system: for example, delaying responses, or pausing in the midst of their utterances, until they had completed a sequence of keystrokes or until a screen change had taken effect. Finally, they found that when patients produced talk whilst the computer was in use, they often began to speak immediately after keystrokes which completed a discrete activity.

To illustrate the conversation analytic approach, we will focus on the last of these three areas: the coordination of patients' unsolicited talk with the doctor's use of the system.

Computer use and the interactional conduct of patients. Doctors and patients stand in different relations to computer systems. The doctors become familiar with the workings of the systems, they are able to enter and retrieve information, and are knowledgeable about the specialised medical information that the computer stores. In contrast, patients have limited access to the systems. They are usually unfamiliar with how systems work and the purposes they serve, and often are unable to view the information which is displayed on the screen. Moreover, patients may not have the specialised knowledge required to make sense of some of the information which is held on the computers and entered and retrieved during the consultation itself.

However, Greatbatch and co-workers' analysis of the video tapes indicated that, despite this apparent asymmetry, the patients attempted to coordinate their own contributions with the doctors use of the system and in some case the operations of the system itself. This was particularly noticeable when patients were initiating an activity – such as seeking clarification concerning the diagnosis or requesting

information about the side-effects of particular treatments – rather than simply responding to doctors' questions. In cases such as this, patients often began to talk after it had become apparent that a doctor had completed a computer-based task and that there was a lull in the use of the computer. Less obviously, patients also often spoke immediately after keystroke sequences, or after individual keystrokes, which completed a discrete activity. This suggested that patients were attempting to synchronise their talk with the doctor's use of the computer.

To explore this possibility, Greatbatch and co-workers located and transcribed every instance in which patients began to speak on their own initiative while a doctor was using the computer. Their transcripts captured not only what was said, but also overlapping talk, pauses within and between the doctors' and patients' utterances, the direction of the parties' gaze, bodily movements such as nods of the head, the positioning of keystrokes in relation to both verbal and non-verbal actions, and the occurrence of changes on the screen. This involved customising and adding to the transcription systems used within conversation analytic research.

Greatbatch and co-workers began by looking at the cases in which patients spoke immediately after keystrokes that completed a keyboard task, such as specifying the form, dosage or quantity of a prescribed drug. Having described the interactional practices used by the patients in these cases, Greatbatch and co-workers then considered deviant cases,[15] that is, cases that did not follow the same pattern. This enabled them to test and refine their initial analysis.

Initial analysis. The fact that patients often spoke immediately after the completion of keyboard tasks suggested that they were able to anticipate potential junctures in the use of the keyboard. Repeated viewing and careful transcription and re-transcription of the video data revealed several aspects of system use that appeared to be involved in this process. It indicated that the patients were synchronising their unsolicited talk with visible and audible aspects of the doctors' conduct, as well as the operation of the system. This involved patients monitoring the doctors' conduct for evidence of the completion of activities. Specifically, patients appeared to be anticipating likely junctures in the use of the system on the basis of the relative intensity of keystrokes, the movement of the doctor's hands and fingers, and shifts in the doctor's orientation and gaze.

In the following example, the doctor informed the patient what treatment he would provide for a penile infection.

Extract 1 [C3:19/ 8/ 91:1.03.40]

Dr: I'll give you cream and tablets.
[(5.2)
[*[Dr typing]*
Dr: [Three times a day the cream
[*[Dr typing]*
[(1.9)
[*[Dr typing; P begins to speak following three successive carriage return keystrokes]*
P: Just [on the outside or
[*[Dr looks at screen as change takes effect]*

A couple of seconds after being told that the cream should be applied three times a day, the patient asked the doctor to explain how it should be applied to the penis. The patient's utterance occurred immediately after a return keystroke, which is the precursor to an extended lull in the use of the system. As the doctor waited for the computer to process information and effect a transformation to the screen-based text, he listened and replied to the patient's query.

Greatbatch and co-workers argued that several aspects of the doctor's conduct, as he makes the three return keystrokes that precede the patient's utterance, projected a lull in system use. First, these keystrokes were louder, made with greater alacrity and, consequently, involved more pronounced hand movements than those which preceded them. Second, as the doctor made these keystrokes with his right hand, the positioning of his left hand did not suggest that he would use it to press a further key immediately after the return keystrokes. Third, the doctor moved his right hand abruptly away from the return key after pressing it for the third time. Upon the completion of the sound of the third return keystroke, the movement and comportment of the doctor's right hand indicated that he would not immediately use it to make another keystroke. Taken together these features suggested potential disengagement from the use of the keyboard and, as it turned out, a lull does follow.

[15] See section 5.3.6.

Deviant case analysis.[16] In the light of detailed analyses such as this, Greatbatch and co-workers concluded that, despite their limited access to the doctors' actions and the operation of the computer, the patients attempted to synchronise their talk with visible and audible aspects of the doctor's use of the computer, as well as of the operation of the computer, such as changes on the screen. To establish the scope and limits of this process, Greatbatch and co-workers analysed deviant cases in which patients' unsolicited speech was not coordinated with junctures in system use. Their analysis of these cases confirmed the importance of visible and audible aspects of system use for the projection and recognition of activity completion. It revealed that even when utterance initiation did not occur immediately after keystrokes that completed a discreet task, the patients continued to display an orientation to the principles identified in the initial analysis.

Thus, for example, the researchers found support for their hypothesis when they considered cases in which patients began to speak during a lull in system use, but not immediately after a final keystroke: i.e. there was a delay. For, in many of these cases the doctors hand and/or bodily movements did not project possible disengagement from the use of the keyboard. Moreover, the initiation of the patients' utterances often occurred after the doctors' conduct had subsequently indicated that a juncture had been reached.

Even patients' utterances which were initiated in the midst of keystroke sequences often turned out to be coordinated with aspects of system operation and use which suggested the possibility of activity completion. For example:

(5a) [HY:2:1 Transcript One]

1 Dr: Okay? [That's great,
[(Dr turns to computer)
2 [(2–0)
[(Dr typing)
3 P: [The only thing other problem I do have u::hm I
[(Dr typing)
4 sleep quite fitfully.

Having completed his utterance, the doctor initiated a sequence of keystrokes. He struck three character keys, the first two with the left hand, the third, centrally-positioned key, with the right. Immediately following the latter, the patient initiated a turn at talk. While this keystroke did not represent the completion of a discrete keyboard task, the doctor's conduct did suggest the possibility of activity completion. Thus, as he struck the third key slightly harder than the preceding two, his left hand was held motionless above the keyboard. Upon releasing the key, his left hand remained still, as he raised his right hand with a pronounced movement (in preparation, it turned out, for a subsequent shift back across to the right of the keyboard). By the end of the sound(s) emitted by the keystroke, the fingers of the doctor's right hand were out of range of the keys; his left hand was motionless. His activities were thus consistent with the completion of a sequence of keystrokes and at least momentary disengagement from use of the keypad.

In summary, Greatbatch and co-workers found that despite patients' limited access to doctor's actions and the operation of the computer, they often attempted to synchronise their talk and visual conduct with the system and its use. Patients monitored visible and audible aspects of the doctor's conduct, as well as of the operation of the computer, to develop a sense of the emerging organisation of the doctor's activities. In this way, patients were able to anticipate upcoming boundaries within the activity and initiate utterances so that they not only avoided interrupting the doctor's use of the system, but in so doing perhaps maximised their chances of securing his or her undivided attention. Moreover, they were able to do this without any clear understanding as to what the doctors were actually doing with the systems at any particular moment.

Having looked in more detail at this and other aspects of the interrelationship between the use of the computer and the interaction between doctor and patient, Greatbatch and co-workers assessed the impact of computerisation on interpersonal communication and professional practice within the consultation. This involved a comparative analysis of recordings made before and after the introduction of the technology.

The impact of computerisation on the consultation. Greatbatch and co-workers' research found that the computer undermined the doctor's ability to document and retrieve information while simultaneously displaying sensitivity to the demands of the interaction with the patient. This was especially apparent in discussions involving topics which

[16] See section 5.3.4 for a discussion of the importance of deviant case analyses.

were tangential to the computational task at hand. Immediately following the installation of the computer this was undoubtedly due in part to the doctors' unfamiliarity with the system. However, analysis of video recordings made more than two years after the installation of the system revealed that its interactional obtrusiveness still compared unfavourably with the use of the paper records and prescription pad and pen.

Greatbatch and co-workers' analysis of the coordination of the participants' interaction with both computer and paper-based documents, identified several properties of the system, which reduced the doctors' ability to participate simultaneously in discussions with the patient while documenting and retrieving medical information. These included the system's ecological immobility, its spatial distribution of the activities of reading and writing, and its reactiveness (see appendix 2).

The research by Greatbatch and co-workers also showed that patients displayed a greater sensitivity to the possibility of interrupting or disrupting an activity in progress when the computer was being used. Thus the coordination of patients' talk with potential junctures in the documentation and retrieval of information was more common when the computer was being used than it was when the doctors were using paper. Greatbatch and co-workers suggested that this was due to the observably interactive character of the doctors' use of the computer; to the ways in which the system observably placed, or was constituted as placing, cognitive and physical demands on the doctor that do not apply when documents are being directly produced by hand or information is being retrieved from the paper records. Thus, the extent to which patients attempted to synchronise their conduct with the visible and audible aspects of system operation and use appeared to depend largely on the ways in which the doctors used the system. For the most part, patients coordinated their actions with the use and operation of the system in situations in which the doctors exhibited a preoccupation with the computational task at hand. The more doctors 'backgrounded' the use of the computer as they interacted with patients, the less likely it was that patients would treat system based conduct as immediately consequential to the production and delivery of their talk.

6.4 Comparison of interactionist and ethnomethodological studies

Qualitative studies differ in terms of whether they focus on participants' perspectives or on participants' practices. Our exemplary studies illustrate the difference. Kaplan's study of the use of a clinical laboratory computer information system ascertained the views of personnel within several laboratories within an academic medical centre. Her analysis revealed that different laboratory workers had different perceptions of their jobs, and that these differences affected how they viewed the system. In contrast, the study by Greatbatch and co-workers of the use of a computerised medical records system within the general practice consultation examined the practices used by doctors and patients before and after the introduction of the system. Their analysis found that the system affected the ways in which participants coordinated and organised their interactions.

The different focus of the two studies is reflected in their sources of data. To ascertain the laboratory workers' perspectives Kaplan gathered data through observations, interviews and open-ended questionnaires. These data enabled her to provide a thick description (Geerty, 1973) of the ways in which the personnel viewed their jobs and the system, and to explain the impact the latter had on aspects of their work. Her research underlines the importance of interactionist research in that it provided insights that eluded the associated quantitative survey, and helped researchers to explain their quantitative findings.

Although Greatbatch and co-workers also used field observations and interviews, they concentrated their analysis on video recordings of consultations before and after the introduction of the computer system. This reflects their interest in practices, as opposed to perspectives. Although ethnomethodological researchers often seek to identify participants' practices through traditional ethnographic methods, a growing number recognise the advantages of using video data. Thus, in the case of the research by Greatbatch and co-workers, the grosser examples of the phenomena they discuss are available from relatively casual inspections of the recordings, and could probably have been noted by a participant observer, or by the participants themselves in the context of *post hoc* interviews and/or questionnaires. However, in order to explicate the processes involved in even relatively simple cases, it was necessary for them to replay the recordings many times, carefully transcribing and re-transcribing the events, and reassessing the emerging analysis in the light of new findings and observations. Moreover, it would have been impossible for them to provide a systematic and detailed analysis of the more complex cases without the use of video, and the possibility of repeated

scrutiny of raw data which it presents. Their study clearly demonstrates that the use of audio-visual data enables researchers to reveal and analyse fine-grained, 'seen-but-unnoticed' aspects of human conduct, which are otherwise unavailable for systematic study. Had they relied solely on data generated through field notes (made whilst observing the scenes) and/or interviews or questionnaires, the tacit, seen-but-unnoticed practices involved in the coordination of organisational tasks and activities would have remained largely, if not wholly, unexplicated.

Nonetheless, Greatbatch and co-workers relied on information gathered through observation and interviews to understand what they were observing when analysing the video data. Without this information their access to the nature and structure of the doctor's activities would have been similar to that of the patients, especially when the doctors were using the computer and/or paper documents. In other words, their analysis would have been restricted to audible and visible aspects of the doctors' conduct. Thus, the study by Greatbatch and co-workers demonstrates how traditional ethnographic inquiry can contribute to the analysis of video recordings.

In spite of their different focus, both studies employed analytic induction.[17] Thus, in contrast to methodologies that involve quantification and/or experimental techniques, they did not begin with *a priori* hypotheses which were then subjected to empirical test. Rather, they used inductive search procedures either to establish what was going on from the points of view of the participants (Kaplan) or to locate recurring patterns of action and interaction (Greatbatch *et al*). This involved an on-going process in which the researchers continued to refine their analysis until an adequate understanding of the phenomena under investigation was reached. In this way, the researchers avoided the imposition of theoretical and conceptual frameworks that had been developed *a priori*. Instead they developed, invoked and refined theories, concepts and analytic categories on the basis of careful observation and description of social phenomena. It was this flexibility and openness which allowed them to provide unique insights into the social organisation of computer use in medical settings.

By using analytic induction (and thus limiting the extent to which the participants' perspectives and practices were refracted through the lens of social scientific theories, methods and concepts), Kaplan and Greatbatch and co-workers sought to maximise the validity of their findings. However, as is often the case in qualitative research, they also employed other methods to validate their research findings. Thus, Greatbatch and co-workers sought to ground their inductive analysis in the orientations and understandings that the participants displayed to each other during their interactions. In the case of Kaplan's study, her findings were validated by two *post hoc* tests: one involving quantitative analysis, the other respondent validation (Kaplan and Maxwell, 1994).[18] The two studies thus illustrate how qualitative researchers draw on a number of different methods to validate their findings.

Finally, the findings of both studies are generalisable beyond the particular settings in which they were developed. As is usually the case with qualitative research, this generalisability depends not on statistical inference but on the development of theories which are applicable in other settings.[19] Thus Kaplan's findings about the relationship between people's perceptions of their jobs and their reactions to computer systems, and the findings by Greatbatch and co-workers on the relationship between social interaction and human computer interaction, provide bases for additional research in a variety of settings.

6.4.1 The practical implications of qualitative research

By focusing on the actor's perspective, interactionism ascertains how the installation of a computer system is or will be viewed by healthcare professionals and patients. By focussing on what people do, ethnomethodology ascertains whether a system supports, undermines or transforms the practices that professionals and patients use and rely upon. Both these types of study can be used to inform the design and evaluation of medical information systems and training programmes concerning the use and impacts of such systems. To date, however, debates surrounding the practical implications of qualitative research in this area have centred on their potential in the fields of design and evaluation.

6.4.2 Evaluating medical information systems

The healthcare community expects that developers of medical information systems will evaluate their

[17] See section 4.6.1.3; [18] Respondent validation has been heavily criticised for failing to take account of the situational factors, which may shape and limit subjects' assessments of social science findings. See section 5.3.1; [19] See section 4.1.1.4.

systems to show that they are safe and effective. At the present time, evaluation in medical infomatics normally involves the use of quantitative methods that have been developed to evaluate drugs and other forms of therapeutic intervention. However, this reliance on a controlled clinical trial (CCT) model of evaluation has recently been criticised for failing "to address the issue of whether users will adopt a system into routine practice" (Forsythe and Buchanan, 1991:8). This criticism has been sparked by the persistent problem of user resistance to systems that have been deemed acceptable by CCT models of evaluation. According to this view, detailed consideration of the social context into which systems are introduced is necessary if the problem of user resistance is to be overcome. This leads to the suggestion that qualitative social research methods should be incorporated into the process of evaluation. Thus, for example, Forsythe and Buchanan argued that:

> the CCT model as it has been adapted to evaluation in medical infomatics is useful, but mainly in relation to the evaluation of system performance. We argue that if system developers broaden their approach to evaluation to include a concern for non-technical; and non-medical issues as well (e.g. users, perceptions of a system), then not only performance issues but also issues germane to acceptance into practice will be examined. In order to accommodate concerns of this sort, however, we will need to extend our methodological repertoire to include other, qualitative methods better suited than the CCT method to collecting and analysing information on social context and subjective experience. (Forsythe and Buchanan, 1992:8.)

Within medical infomatics the interest in qualitative methods is focused largely on interactionist approaches. It is suggested that participant observation and interviewing should be used to discern the perceptions of users and, although reference is also made to the need to understand patterns of working and the social contexts into which systems are introduced, these phenomena are approached, as in Kaplan's study, from an interactionist position (Fafchamps, 1991; Fafchamps, Young and Tang, 1991; Werner and Schoepfle, 1987; Zubroff, 1988). However, the studies by Greatbatch and co-workers suggest that ethnomethodological research could also make a useful contribution to the evaluation of computer systems by identifying how their use impacts upon participants' practices.

It is important to note that, although interactionist and ethnomethodological studies chart the effects of computer systems, they do not establish whether or not these effects are desirable. Such judgements depend upon the interests, objectives and opinions of organisational stakeholders and decision makers and, consequently, the same results may be interpreted in different ways by different groups. The primary objective of evaluation studies that use qualitative research methods is to provide information to interested parties and decision makers (Anderson *et al,* 1994b), rather than to pass judgements themselves.

Those who advocate the use of qualitative methods in medical infomatics have suggested that qualitative investigations should be incorporated into the design process, so that systems can be configured from the outset in ways which are sensitive to the perspectives and working patterns of users. To date, little progress has been made in this direction, though studies such as Kaplan's do give some indication of the types of information that interactionist studies might make available during the early stages of design. However, greater strides have been made toward using qualitative methods to inform the design of systems within the field of requirements analysis. Interestingly, however, the emphasis has been on the use of ethnomethodological approaches.

6.4.3 Using qualitative methods to specify requirements for computer systems

The past few years have witnessed a growing dissatisfaction with traditional approaches to requirements analysis; that is, "the process of extracting information about the required functionality and other properties of [a computer] system" (Jirotka and Goguen, 1994:20). This dissatisfaction has been fuelled, in part, by a series of high profile failures of computer-based information systems, such as the Computer Aided Dispatch system which was installed by the London Ambulance Service in 1992. These failures, together with less highly publicised problems, have given rise to questions concerning the adequacy of conventional methods for specifying requirements for computer systems. In particular, it is suggested that these methods are insensitive to the social organisation of work and interaction in the settings into which technologies are introduced. According to this view, a much more detailed understanding of the work and interactions of those who will use a particular system is essential if unanticipated and in many cases unwanted consequences are to be avoided when the system is deployed.

This turn to the social has led to an interest in the computer sciences in the possibility of using qualitative social research to specify requirements for new technologies. Most of the exploratory

work in this area has advocated the used of ethnomethodological approaches, and thus has focused on participants' practices rather than participants' perspectives. This ethnomethodologically oriented work, which involves the use of both traditional modes of ethnographic inquiry and video analysis, confirms that, despite recent innovations, many of the methods used within requirements analysis are based on inadequate conceptualisations and understandings of the ways in which people accomplish and coordinate organisational tasks and activities (e.g. Luff *et al,* 1994; Suchman, 1987). Although, there is not yet an established method for transforming the insights of such studies into requirements than can be used to inform the design of systems, significant advances have been made towards achieving this goal by a number of research teams consisting of both social and computer scientists. These teams have started to develop guidelines for collecting and analysing both ethnographic and video data, and for using the findings to inform the specification of requirements (Heath and Luff, 1992a, 1992b; Hughes *et al,* 1994).

In the field of medicine, Greatbatch and co-workers' study of the use of computers in general practice illustrates how ethnomethodological studies could be useful for specifying requirements for healthcare systems (Luff et al, 1994). In the light of their findings, Greatbatch and co-workers formulated a preliminary set of requirements for computerised record systems to support social interaction within the general practice consultation. Then, having concluded that these requirements could not be satisfied through extensions of current systems, they reviewed a range of more advanced technologies, which could possibly be used to develop innovative systems. These included scanner technologies, hybrid systems which combine paper and electronic documents, and mobile devices. They concluded that none of these technologies would satisfy the initial requirements completely. However, their assessment highlighted the various trade-offs that could be made in developing an innovative system for general medical practice.

The possibility that ethnomethodological research might make an important contribution to the specification of requirements for computer-based information systems has also been explored in the context of financial dealing rooms, urban transport control rooms and the offices of an international news agency. However, although these studies suggest that this approach could make an important contribution to requirements analysis and design, they also highlight a number of potential obstacles to the transformation of the analytic frameworks used by ethnomethodological researchers into an applied method. To begin with, ethnomethodological studies, especially those involving the use of video data, often involve extremely detailed analysis. However, it is not clear what level of detail is necessary for the practical requirements of software engineers and system designers. In addition, there is the related question of whether it will be possible to develop a 'quick and dirty' method, which would shorten the time currently required to complete the analysis, whilst capturing the detail necessary in order to make a distinctive and useful contribution to the requirements process.

The same problems would arise in relation to interactionist research. To date, this approach has not figured much in the debates surrounding the role of qualitative social research within requirements analysis. Nonetheless, it is consistent with requirements methods which seek to incorporate the views of users into the design process. Consequently, interactionism could make a contribution to these methods, or might even form the basis for a new method of capturing requirements for computer systems.[20]

6.5 Conclusion

The two strands of qualitative research on information systems involve alternative forms of social analysis, with different interests and objectives. However, they both provide access to social processes that are not amenable to quantitative analysis, and generate findings which can be used to assess the likely organisational impacts of systems that are available to purchase, about to be implemented, or which are under development.

6.6 Links between the case study and the full report

Taken together, the two studies examined here illustrate the following aspects of qualitative research as applied to HTA.

[20] See Fafchamps (1991) for a discussion of how the findings of a qualitative study, which centred on the perspectives of two groups of physicians, were used to inform the design of a physician's workstation.

- Qualitative and quantitative research are complementary and can be combined in the same study.[21]
- Qualitative research involves a constant interplay between induction and deduction as patterns identified through induction are then tested deductively.[22]
- There are some problems in HTA for which qualitative methods provide the technically superior option.[23]
- Qualitative research can usefully take theoretical models that have been developed in other fields and examine their power in relation to HTA empirically. Such research can, in turn, contribute to cumulative knowledge.
- Qualitative research can provide a rigorous descriptive base which, among other things, alerts policy makers to aspects of the situation that can influence outcomes significantly but remain undetected by other methods.[24] The flexibility of qualitative research design permits the researcher to identify significant but unanticipated factors and to include these in subsequent analyses.[25]
- Qualitative methods may be used in the developmental stages of research to clarify the research question, aid conceptualisation and generate hypotheses for later research.
- However, qualitative methods can also be used to test hypotheses.
- Qualitative research is useful in interpreting, challenging, illuminating and qualifying the results from quantitative research.
- Qualitative research makes an important contribution to our understanding of the ways in which contextual factors and local contingencies impact upon the outcomes of technologies.[26] This improves the both the explanatory power and the generalisability of such research.
- Qualitative research allows us to examine the process by which technologies are implemented and to consider the implications of such processes for outcome.[27] In this way they can uncover the causal mechanisms that link intervention and outcome, allowing greater understanding of how modifications of such processes could improve outcome.
- The range of methods includes participant observation,[28] interviews,[29] documentary analysis,[30] and CA,[31] which are employed, often alongside one another, in qualitative research.
- Qualitative research involves the use of the logics of the constant comparative method,[32] analytic induction,[33] and deviant analysis.

[21] See section 3.1.1; [22] See section 3.2.1.2; [23] See section 3.1.1; [24] See section 3.2.2.2; [25] See section 3.2.2.5; [26] See section 3.2.2.3; [27] See section 3.2.2.4; [28] See section 4.2; [29] See section 4.3; [30] See section 4.4; [31] See section 4.5; [32] See section 4.6.1.4; [33] See section 4.6.1.3.

Chapter 7

Qualitative methods in programme evaluation

> Evaluation entails a view of society. People differ about evaluation because they differ about what society is and what it ought to be. Much of the debate about evaluation is ideology disguised as technology. (Hamilton *et al,* 1977:25.)

Many published evaluations cite no methodological references (e.g. Lord and Green, 1995; Speirs and Jewell, 1995; Gordon *et al,* 1997), and so this case study is presented as a guide to the literature of that "exciting" (Shortell and Richardson, 1978:xi) area with particular emphasis on the place of qualitative methods of inquiry.

The original description of this part of the review was 'the evaluation of organisational innovation', from the subtitle to Smith and Cantley (1985), but this has been amended to 'programme evaluation' to reflect the fact that the literature produced by a search oriented to evaluation tended to focus on its use in the study of specific programmes, rather than of the organisations in which they were located. (Examples of the use of qualitative research in organisational communication can be found in Herndon and Kreps, 1995).

Much of the material for this section has been assembled by pursuing references cited in the bibliographies in a personal collection of texts and papers acquired during the course of teaching classes in qualitative methods and conducting two evaluation projects. The first project has already been mentioned. The second was a more orderly study relating to community pharmacists and the care of people with mental health problems. Both were primarily qualitative in nature.

Searches of computerised databases (Sage, Social Sciences Citation Index – terms: evaluation research, qualitative evaluation) were done in parallel with those carried out for the main body of this work but added little additional material, and more significantly, omitted key publications. Computerised databases may be of limited value in tracking methodological issues, as others have noted. In Filstead's (1981:260) words:

> Computerised literature searches, while yielding a number of key sources, were not as helpful as first thought because the term is not free of ambiguity, and there are a variety of meanings to the term qualitative as well as other terms that cover the same general area (e.g. participant observation, ethnography, field research and so forth).

Krantz (1995:89) noted that someone relying on this method would find a set of citations biased towards set-piece public confrontations rather than describing the working practices of real researchers doing real studies.

In section 7.1, the literature on the use of qualitative methods in programme evaluation is reviewed. Section 7.2 consists of an examination of the research methods used in a sample of evaluation studies published in the 1995 issues of seven selected journals.

7.1 The use of qualitative methods in programme evaluation

7.1.1 The requirement to evaluate

Evaluation is increasingly seen as an essential part of any intervention or programme of action in health, education and social services; not to evaluate equates with irresponsibility (Smith and Cantley, 1985; Stake, 1986; Walker, 1993; Reinharz, 1994). Scriven (1991) noted, however, that evaluations may be conducted on products and people as well as on programmes. Within the field of health care much of the emphasis on evaluation has been related to the introduction of new technology (Jennett, 1992; Smith, 1995).

Coulter (1991) has described a hierarchy of evaluations. Level 1 is the evaluation of specific treatments; level 2 evaluations examine the patterns of care for particular patient groups; level 3, the evaluation of organisations; and level 4, the evaluation of health systems. Most programme evaluations seem to span levels 2 and 3. Although Scriven (1991) argued that the logic of evaluation is the same across different forms and objects, this was contested by Greene (1994). She proposed that they constitute radically different tasks requiring responses which are fundamentally different in character. Similar arguments could be made about the different levels of evaluation, though they are not always

presented in an explicit form. For example, although Cochrane (1972:1) talked about evaluating the NHS, his actual examples are evaluations, or failures to evaluate, in respect of specific 'interventions' and discrete treatments.

As Shortell and Richardson (1978) pointed out, though, there is also a need to understand the political and administrative environment in which research and evaluation dealing with human service programmes are likely to occur. 'Better' evaluation may not be simply a matter of improved design or more precise measurement but should include some acknowledgement of political nature of evaluation itself. Programmes are suggested, debated, accepted, prioritised, implemented and usually funded through political processes and, in contrast with many other forms of research, evaluators have to decide whose questions are to be addressed and whose interests are to be served by their work (Greene, 1994). Managers may not view this aspect of evaluation as problematic and their requests for 'evaluations' will almost invariably be framed in terms of the management needs of the system. Evaluation research is now regarded as a key management tool. (Freeman and Solomon, 1981; Ong, 1993). However, changes in patterns of care through the introduction of any particular programme will almost inevitably have organisational consequences. Commonly such organisational innovations require fundamental modifications in professional alignments and routines and may threaten the roles, status and even the economic security of particular individuals or groups (Mechanic, 1978; Shortell and Richardson, 1978). They may also call for changes in the attitudes and behaviour of service users, who may mobilise external lobbies – self help groups, local politicians, etc. – to influence the nature of evaluation and the changes that it is possible to introduce. A successful evaluation is likely to require some recognition of the interests of all of these groups or 'stakeholders' if a manager is not to be presented with proposals that may seem rational but which cannot be implemented.[1] Weiss (1984:256) interpreted the word stakeholders to mean either the members of the groups palpably affected by the programme and who will conceivably be affected by evaluative conclusions about it or the members of groups that make decisions about the future of the programme.

Evaluation research does not differ from basic research in the methods it employs. Instead, its agenda is determined not by the advancement of science or of social theory but by what commissioning stakeholders expect to find useful (Kiresuk *et al,* 1981; Berk, 1995; Stake, 1986). It has different purposes and different audiences. The evaluator's task is not that "of determining truth for the ages, but the best possible advice at the time it is needed" (Scriven, 1991:153).

7.1.2 Programme evaluation – a case history

Case studies frequently begin with a case history. So do many evaluations – Cronbach *et al,* (1980:7) even refer to the evaluator as "essentially an historian". Many evaluation texts and articles, regardless of specific subject area, start with a broad history of the field (e.g. Robinson, 1982; Smith and Cantley, 1985; Shadish *et al,* 1991; Marsden and Gissane, 1992; Whiteley, 1992). Much of this history is situated in the literature of educational evaluation and comes mainly from USA sources. It is not necessary to explore this in detail but it is important to recognise that this is where much evaluation experience, terminology and theory has originated.

> In terms of world culture, perhaps the two most distinctive US contributions of the 1970s are the movie *Star Wars* and evaluation research. (Freeman and Solomon, 1981)

Evaluation research, which should be distinguished from evaluative research as described by Black (1995), became identifiable as a discrete area of study in the 1960s. In the USA this was a period of optimism about progressive social policy and also about the ability of social science to inform the policy process (Hamilton *et al,* 1977; Rossi, 1981; Bulmer, 1987). The welfare state characteristic of Northern Europe had never developed there but the Kennedy/Johnson years saw the introduction of numerous social, educational and healthcare programmes intended mainly to help the underprivileged (Albaek, 1995). Levine (1981) has pointed out why the USA, in contrast with other western nations, implements its policies through programmes: the social and political/constitutional structures are markedly different and the USA has both a greater tendency, and it is claimed, a greater need, to identify population subgroups and direct programmes at them. In order to obtain political support for public expenditure in these specific ways, however, the sponsors had to be able to show that the money allocated had been spent

[1] See section 5.3.7 for a discussion of fair dealing in research.

effectively, where effectiveness was synonymous with scientifically guided action. The dominant social science paradigms in the USA are rationalist and positivist and not surprisingly these became the main paradigms of evaluation research during that period. As a result the field was dominated by input–output studies using survey or attitude measurement tools, which attempted to establish quantitative descriptions of programme effects that could feed into various measures of statistical association in order to identify causal direction and significance. To the extent that other nations have followed American models of social policy, they have experienced a similar movement towards programme-based provision and programme evaluation.

Evaluation became an independent and expanding industry, forming professional organisations in the early 1970s: evaluation research was professionalised (Rossi, 1981). It also produced a voluminous (and self-referential) literature, much of which relates to programme evaluation, the largest area of evaluation to which a self-conscious specialty has been devoted. Within this the three most active sub-areas are education, health and law enforcement (Scriven, 1991:285).

In describing the foundations of programme evaluation and the place of seven key theorists within it, Shadish and co-workers (1991) documented and analysed the changes in evaluation theory. As evaluators began to evaluate evaluations, it became clear that this scientific testing of programme effectiveness had turned out to be far more difficult than anticipated. The organisations and programmes studied simply did not fit the evaluative models used. Moreover, the evaluation findings were rarely used (Rich, 1981; Patton, 1986).

In summary:

- the assumptions upon which this rationalist model of evaluation were based came to be seen to be flawed (see below in the discussion of the British debates)
- the studies were very expensive to perform and frequently the results were neither timely nor useful.

In consequence evaluation theories and methods changed to accommodate other methods. Greene (1994) suggested that this resulted from the combined force of the political-contextual and methodological-philosophical arguments which "catalysed the development and later acceptance of a diverse range of alternative approaches to program evaluation".

7.1.3 Qualitative methods – adoption and apostles

> I can't tell you how we had to battle to do case studies. And I remember people – not you or any of your friends, of course – who said – 'Don't tell us about this process stuff, you know, these case studies, this qualitative data, yuk.' And the only way we were able in 1973 to justify doing the casework was to tell them that we would use the case studies to validate the survey, and that was fine. (McLaughlin M cited in Alkin, 1990:38.)

Shadish (1995) considered the introduction of qualitative methods into the repertoire of tools as one of the milestone accomplishments of programme evaluation. They may be used as an adjunct to quantitative methods or are sometimes the main or sole methodology.[2] Major figures usually identified with experimental or scientific approach to evaluation began to note the potential for qualitative methods in evaluation from the early 1970s. Campbell (1978), co-author of the influential *Experimental and quasi-experimental designs for research* (see Shadish and Epstein, 1987 for a measure of this influence), wrote of his own ambivalence with regard to case studies. At the same time he made an overwhelming argument for the recognition of qualitative knowing in action research – his then preferred name for programme evaluation – and also in laboratory research. He appeared to accept qualitative and quantitative evaluations as alternative forms of inquiry:

> If qualitative and quantitative evaluations were to be organised on the same programmes, I would expect them to agree. If they did not, I feel we should regard it possible that the quantitative was the one in error. (Campbell, 1978:200.)

However in this, and other writings, he preferred to stress the predominantly complementary or contextual function of qualitative methods:

> In program evaluation the details of program implementation history, the site-specific wisdom and the gossip about where the bodies are buried are all essential to interpreting the quantitative data. (Campbell, 1984:30.)

[2] See section 3.1.1.

Shadish and co-workers (1991) likened Campbell's analysis of the relationship between quantitative and qualitative research to his analysis of the relationship between experiments and quasi-experiments. "In each case he constructs a case of **relative** superiority; but he produces such a strong case for the inferior alternative and such a qualified case for the superior option that the marginal difference between the superior and the inferior **seems** minor". Other readers may give different weights to Campbell's criteria for judging method choices – if they prefer to know more about the range of variables which are relevant rather than more about the causal relationships between them, then qualitative methods may rank higher.

Despite the assertion (Shadish, 1995) that the introduction of qualitative methods was slow, and controversial, a bibliography of their use in evaluation research was produced in 1981 (Filstead, 1981). Filstead noted how evaluators, in addition to policy makers, were realising that the natural science (quantification) model of evaluation lacked "the ability to tap the contextual understandings about the processes and understandings involved in social interventions".

Writers currently most immediately identified (and most frequently cited) with the use of qualitative methods in the field of evaluation are Patton and Guba and Lincoln, though, importantly, none of them dismiss quantitative methods. Guba and Lincoln (1989), in their *Fourth Generation Evaluation,* trace the origins of contemporary evaluation through its first three generations. Two of these pre-dated evaluation research. The first generation was contemporary with the development of scientific management informed by the spirit of Taylorism and its application in companies like Ford. It is exemplified by the **measurement** of various attributes of schoolchildren, for example, in early IQ tests, which apparently produced very precise quantitative data. In effect, measurement and evaluation were used interchangeably – a phenomenon which is still evident in league tables of school and hospital effectiveness.

First-generation studies tended to see measurement as an end in itself. However, by World War II there was growing interest in the reform and modernisation of educational curricula. Researchers became more interested in specifying objectives for children's learning and comparing the effectiveness of different methods of achieving them. Guba and Lincoln (1989) have termed this the second generation of evaluation, an approach characterised by the description of patterns of strengths and weaknesses in the attainment of these stated objectives. With experience, programme developers became unhappy with the rigidity of this approach. The need rigorously to pursue previously defined objectives made it impossible to react creatively to successes and failures that emerged as a programme was being delivered. Evaluations that only deliver-ed their findings after the completion of the programme were too late to be of much use in improving the programme as it happened. Finally, there was growing dissatisfaction with the exclusion of issues of judgement: did the programme have the right objectives operationally defined in the right way to start with? Guba and Lincoln presented third-generation evaluation as the response. This treated objectives as problematic, evaluating goals as well as performance. Judgement, of course, implies the existence of standards and here values intrude into a supposedly value-free enterprise.

Guba and Lincoln (1989) suggested that all of these previous generations of evaluation are vulnerable to three main criticisms: a tendency towards managerialism, a failure to accommodate value pluralism and an over commitment to scientific (as opposed to humanistic) modes of inquiry. Almost inevitably the result is overdependence on formal quantitative measurement. They described fourth-generation evaluation as an alternative approach – responsive and constructivist.[3] Such evaluations are responsive in that they seek out different stakeholder views, which are incorporated into subsequent data collection. (Stakeholders are categorised as **agents**, **beneficiaries** and **victims**.) Fourth-generation studies adopt a constructivist paradigm; a paradigm which is also referred to as naturalistic, interpretive and hermeneutic, or possibly (mistakenly) as qualitative (Guba and Lincoln, 1989:45). The final phase of the evaluation includes the preparation of an agenda for negotiation to reach conclusions jointly agreed by the main stakeholders. This may indeed be fourth generation but it is clearly prefigured by MacDonald's evaluator as the collector of judgements (quoted in Walker, 1993:140), the incorporation of participant judgements (Campbell, 1978) and the pluralistic and constructivist model of evaluation suggested by Smith and Cantley (1985).

Fourth-generation evaluation makes fairly regular[4] appearances among methodological citations; though this may demonstrate the utility and relevance of its methods rather than a full endorsement

[3] See sections 3.12 and 3.2.1.1; [4] See section 3.2.1.1.

of the underlying paradigm. While Heap (1995) noted the merits of Guba and Lincoln's (1989) methods and praised their commitment to democratic ideals, he also showed that the paradigm is not as coherent as it might seem. In particular, he suggests that Guba and Lincoln have confused two rather different versions of constructivism. They tend towards the radically subjectivist position known as **cognitive constructionism**, where everything is thought to be constructed in the mind much as Berkeley imagined.[5] However, this is an unnecessarily extreme position, which undercuts the notion of evaluation itself. How can you evaluate anything if everything is subjective? Heap argues for a shift towards the other tendency in Lincoln and Guba's work, social constructivism, where the world is treated as a product of social interaction which can be observed and described.[6]

7.1.4 The quantitative–qualitative debate

Methodological prejudice has been a major concern in evaluation. This is reflected in attempts to debate the absolute rather than the relative merits of quantitative and qualitative methods, mirroring arguments within the social sciences more generally (Patton, 1986).[7] Shadish, (1995:47) also pointed out that this is a contest not confined to the field of evaluation and one which has many dimensions. As Campbell (1978; 1979) explained, the terms quantitative and qualitative are also used as shorthand for a number of related and overlapping concepts, for example, scientific, *naturwissenschaftlich* versus humanistic, *geisteswissenschaftlich*.[8] There are obvious parallels within the field of medical practice – for example, Engel's (1977) call for a biopsychosocial model to replace that based on biomedicine. It is likely that certain aspects of the debate would be better understood as part of the process of professionalisation of both social science and evaluation research (Rich, 1981:180; Rossi, 1981). Many evaluators were/are psychologists (Shadish and Epstein, 1987). Psychology has modelled itself more closely on the physical sciences than have most other social sciences and, in both the USA and the UK, has been the most active professionaliser. The claim to an absolute scientific foundation sustains psychology's demand for the kind of legally backed licensing regime that older professions have established, allowing it to exclude potential competitors from important sections of the market for applied social science.

The inaugural volume of the *Sage Research Progress Series in Evaluation* (Cook and Reichardt, 1979) was devoted to the quantitative–qualitative debate as, 16 years later, was a special feature of the journal *Evaluation And Program Planning* (Shadish, 1995). Krantz (1995) in one of the articles in this feature reproduces a table from Reichardt and Cook (1979), which shows the attributes of the qualitative and quantitative research and evaluation paradigms, and demonstrates how closely these resemble the attributes of the romantic and empiricist paradigms *(Tables 1* and *2*).

See printed copy for Table 1

[5] See section 2.2.3; [6] See section 3.2.1; [7] See section 3.2; [8] See section 2.3.1.

TABLE 2 *Attributes of romantic and empiricistic paradigms. Reproduced from* Evaluation and Program Planning *Volume 18, Krantz DL. Sustaining vs. resolving the quantitative–qualitative debate. p. 89–96. © 1995 with permission from Elsevier Science*

Romanticism	**Empiricism**
Person-centred	Environment-centred
Individual creative	Individual reactive
Behaviour is purposive	Behaviour is reactive
Man – inherently good	Man – inherently neutral or bad
Social system constrains individual's nature	Social system defines individual
Reality – indeterminate, complex, mysterious, constructed	Reality – determinate, discoverable
Behaviour is free, spontaneous	Behaviour is controlled
Reality is apprehended, knowable through intuition	Reality is investigated by observation and rational thought

Given that the conflict between romanticism and empiricism goes back at least 2000 years,[9] it is perhaps not surprising that the problems have not been resolved in the last 30 years:

> Philosophers have come to realise that there is no 'right' position, nor can some ready amalgam be made. Rather, they represent two postures or attitudes to knowing that are rooted in stylistic preference rather than evidence or argument. (Krantz, 1995:90.)

Research methods can be chosen without necessarily involving philosophical considerations as Reichardt and Cook (1979) showed. Indeed Greene (1994) noted how many evaluation practitioners and audiences missed the quantitative–qualitative debate entirely. Evaluation research could, and maybe should, be regarded as methodologically opportunistic, the procedures being dictated by the empirical problems rather than by adherence to some particular school of thought.[10] In Scriven's terms: "Evaluation is a pragmatic discipline, and if one can't grasp that and implement the consequences, one isn't very bright". (Scriven, 1991:211). Nevertheless, it is important to be aware of the extent to which a technical choice of methods also involves an implicit choice of methodology and, hence, a philosophical position.

7.1.5 The British debate

Russell (1996) characterised contemporary approaches to evaluation in the area of health care as a debate between the followers of the evangelist Cochrane and those adopting the broader approach proposed by Illsley (1980).

However, it may be better to see this as a contrast between people influenced by two different generations of thought on evaluation, and using examples from different disciplinary areas. Cochrane was an epidemiologist writing at the time of highest hopes for quantitative methods, though even he conceded that they are not appropriate for all questions in HSR. Illsley was writing as a sociologist, influenced by the **illuminative** evaluation model proposed by Parlett and Hamilton (1977). By 1980, evaluation was looking a much more difficult task than Cochrane could have imagined. Smith and Cantley (1985) developed Illsley's position in a way which drew more explicitly on US experience. They followed Illsley's account of the main features of the classic evaluative model from the 1960s – roughly equivalent to Lincoln and Guba's second generation:

- the primary objective of the intervention can be unequivocally specified
- it tests the effectiveness and/or efficiency of a given product or process in achieving the goal compared with alternative interventions or with no intervention
- it has a precise, foreseeable and measurable control over the nature and quality of the input
- influences extraneous to the measured input, the controlled intervention process and the measured output can be excluded by the research design
- the criterion of success is uncontroversial and can be measured on a single dimension.

[9] See chapter 2; [10] See section 3.1.1.2.

Smith and Cantley argued that three assumptions underlie such an approach to evaluation: the first is that of **rationality**. It is assumed that policy making in health and welfare agencies is a rational process with clear goals. In reality, policy making is a political process where the result is usually a compromise between different interests with goals that are multiple, complex, frequently conflicting and likely to vary over time. Even if there is agreement over what these might be, in many circumstances valid output measures are not available. A rational model is also said to make a clear distinction between means and ends – between process and outcome. Others, including Coulter (1991), have thought that this was a somewhat artificial distinction.

The second assumption examined by Smith and Cantley is the desirability of **experimental design**, epitomised by the RCT. For many drugs or clinical procedures this is probably the optimal approach, though even here there are problems (Black, 1996). As Illsley pointed out, however, in service evaluations, it is often impossible to achieve any or all of the necessary controls over goals, inputs and output measurements. Even Cochrane (1972:3) conceded that his emphasis on RCTs has limited applicability to what he calls "board and lodging and tender, loving care", by which we might understand much of the current domain of HSR. The RCT is also of limited value in explaining **why** and **how** any change has occurred.

According to Smith and Cantley, the third assumption frequently made is that of consensus in a professional organisation. To equate the interests of an organisation as a whole with goals decided by its senior managers is politically naive. More often than not there are multiple participants with multiple views, as stakeholder approaches recognise.

If the conventional assumptions are adopted, then quantitative approaches would seem appropriate to most of the questions asked in HTA. This model can work when we have a rational specification of goals in an environment where all relevant variables can be identified and controlled and where there is a consensus that makes the assignment of numerical values to the variables uncontestable. The nearer that this model is approached, the more successful a purely quantitative approach will be, as in drug trials. Conversely, the further the actual research situation departs from this model, the less appropriate it becomes. This can be expressed in a slightly different way: the choice between mainly quantitative and mainly qualitative forms of evaluation is likely to rest on the degree of knowledge of the phenomenon to be studied. The more that is known about a programme and its underlying theories, the more possible and logical it becomes to use experimental design. The less that is known, the more it makes sense simply to try to understand the basic components.

The evaluation Smith and Cantley (1985) undertook was that of a new psychiatric day hospital. They described two research objectives. The first was to provide an account of its services and to suggest what factors were most influential in the course of this development. The second was more general; they aimed to explore some of the conceptual and methodological problems of evaluation with particular reference to defining objectives, measuring outcomes and determining criteria of success. Data were collected from hospital records, field notes on observations and conversations, recordings of meetings involving relatives and staff and interviews with staff members and relatives. From the interviews six different ways of defining success were identified – free patient flow, clinical cure, provision of an integrated service, beneficial effect on related services, support for relatives and a high-quality service. The organisation and analysis of data followed these categories. The conclusions were similarly grouped around these themes.

Smith and Cantley's call for political and methodological pluralism in evaluation research, partially echoes the scientific pluralism advocated by Kinston (1983) in relation to the whole HSR enterprise, and the general direction of the US evaluation writings discussed above. This includes, but also extends, the quantitative–qualitative debate.

7.1.6 Current practice in evaluation research

Having provided a brief overview to the field the intention is now to examine some aspects of current practice with particular reference to definition, types, models and designs. All have implications for the choice of method(s).

Defining evaluation. It may seem odd to have arrived mid-way through this section without giving a definition of evaluation. Guba and Lincoln (1989:21) declared that "There is no answer to the question, 'But what is evaluation really?' and there is no point in asking it". However, their history of evaluation outlined (at some length) above indicates how at various times the word evaluation has been associated with various technical and, later, moral/political value judgements. The problem of definition is further complicated by the fact that

evaluation is a commonplace personal, as well as an institutional, activity (Pearsol, 1985; Stake, 1986), and even within the professional sphere it is carried out at various levels ranging from everyday record keeping to an external report (Scott, 1992:61).

Most writers do provide a definition – the difficulty is however to know whether the preferred mode of inquiry precedes or influences the choice of definition or vice versa. The point is, of course, that different definitions reveal important differences in what various evaluators emphasise in their work, for example, the questions they wish to address (or exclude) and the preferred methods for answering them.

Texts written from the epidemiological/public health standpoint tend to stress measurement, rationality and the mainstream evaluative model:

> Evaluation is based mainly on rational thinking and on measurable data. It is not initiated to **prove** a particular point of view but with the aim of **improving** a particular activity. (Holland, 1983:xv).

Although the evaluation of health education (an integral, if unquantifiable, part of any primary healthcare service), "calls for different methods, probably relying more on anthropology and sociology" (Holland, 1983:xv).

St Leger and colleagues also discuss the use of evaluation in the context of health services:

> The term 'evaluate' has in general usage been defined as 'ascertain the amount of; find numerical expression for'. In the context of a health service, we extend this notion to define evaluation as: The critical assessment, on as objective a basis as possible, of the degree to which entire services or their component parts fulfil stated goals. (St Leger *et al,* 1992:1.)

These definitions can be compared with those offered by Scriven in his *Evaluation thesaurus* (1991:vii): "This is a book about evaluation in the everyday sense in which it refers to the process of determining merit, worth, or value of things – or result of that process" (programme evaluation is one sub-area of this effort),

or by Cronbach and associates:

> By the term **evaluation** we mean systematic examination of events occurring in and consequent on a contemporary program – an examination conducted to assist in improving this program and other programs having the same general purpose. By the term **program** we mean a standing arrangement that provides for a social service. (Cronbach *et al,* 1980:15.)

At first sight it does not seem possible to reconcile these different approaches, though the definition (and practice) proposed by the WHO "draws explicitly and intelligently on the developing general literature and practice of evaluation" (Marsden and Gissane, 1992).

> Evaluation is a systematic way of learning from experience and using the lessons learned to improve current activities and promote better planning by careful selection of alternatives for future action. This involves a critical analysis of different aspects of the development and implementation of a programme, its relevance, its formulation, its efficiency and effectiveness, its costs and its acceptance by all parties involved. (WHO, 1981:11.)

On the other hand, this might be compared to the WHO definition of health as being so all-encompassing as to be of little practical use to the commissioner trying to decide what approach to purchase for what purpose.

As these quotations suggest, evaluations have different uses. An instrumental use, in the terms of Holland (1983) and St Leger and co-workers (1992), suggests some direct linkage and application of the evaluation findings and recommendations to policy making and programme planning. A conceptual use, which is more the WHO sense, suggests an indirect application whereby policy makers and planners begin to think differently about the problem, or frame policy approaches in a new manner (Rist, 1995:xiii). (Rossi and Freeman, 1993 add a third type of use – persuasive evaluation – where existing results are enlisted in efforts to support or attack political positions.)

7.1.7 Evaluation designs, models and types

7.1.7.1 Evaluation design

Drawing on earlier work by Cronbach, Rossi and Freeman (1993:31) described any evaluation as an art. "The design (the plan for allocating investigative resources) must be chosen afresh in each new undertaking, and the choices to be made are almost innumerable ... For any evaluation many good designs can be proposed but no perfect ones" Prout (1992:77) and Walker (1993:140) also stressed the idiosyncratic nature of programme evaluations. From Walker: "It is reasonable to evaluate this programme with this staff, these resources and in these circumstances. Evaluation by definition is about specifics, never about generalities". Its methods may be derived from other disciplines – but these are selected and synthesised in a way that best serves some particular evaluation.

7.1.7.2 Models of evaluation

The literature offers a number of models of evaluation which reflect various assumptions about the nature of the task and the relationship between the evaluator and the stakeholders. These models are ways of framing evaluations rather than specific blueprints. Some of those most closely associated with qualitative methods are:

- **goal-free evaluation**: gathering data on actual effects and evaluating their importance in meeting demonstrated needs, without discussion of goals, therefore avoiding the possibility of missing unanticipated effects (a Scriven model described in Patton, 1990)
- **responsive evaluation**: emphasising continued contact with programme staff and dealing with issues as they arise (Stake's model as described by Guba and Lincoln 1989, Patton, 1990 and Scriven, 1991)
- **illuminative evaluation**: concerned with description and interpretation rather than measurement and prediction and seeking to explore a vast array of questions (Parlett and Hamilton, 1977)
- **pluralistic evaluation**: informed by theories of political pluralism and sensitive to the ways in which different groups define success (Smith and Cantley, 1985)
- **fourth-generation evaluation**: see discussion of Guba and Lincoln (1989) above.

7.1.7.3 Types of evaluation

Process and outcome evaluations. Outcome evaluations attempt to assess the effects produced by programmes or policies. Outcomes may be immediate, end of treatment or long term – some may be short-lived, others persist. A process evaluation focuses on the variables between input and output and aims at illuminating and understanding the internal dynamics of a programme.[11] Process evaluations investigate informal patterns and unexpected consequences as well as formal activities and anticipated outcomes.

Formative and summative evaluations. The distinction between studies that ask how good a service is and those that ask how the service can be improved has been around for many years. To polarise the discussion Scriven introduced the terms formative and summative (Cronbach *et al,* 1980). A formative evaluation is typically conducted during the development or improvement of a programme. It is conducted with, and for, the staff involved with the intention to improve the programme. Often this is reliant on site visits, direct observations, surveys and in-depth interviews. Both Mechanic (1978) and Shortell and Richardson (1978) stressed the need for these in the field of health care. A summative evaluation is conducted after the completion of a programme or, for a continuing programme after it has stabilised, and frequently for the benefit of some external audience. According to Scriven (1991) this should not be confused with an outcome evaluation. Both process and outcome evaluations may, in principle, be either summative or formative in character.

For any evaluation the distinction between the summative and formative elements is critical but as Cronbach and co-workers (1980) pointed out almost any evaluation when **used** will have a formative function.

Impact evaluation. This attempts to assess the longer term and more general results of programme operations. Ong (1993) regards this as the most difficult and often avoided, aspect of evaluation research. Ormala (1994) has discussed the European experience of qualitative methods and practice in the evaluation of government science, technology and innovation policies and programmes. Not all the dimensions of impact identified are of direct relevance to healthcare activities but others are – the individual and organisational learning effects and behavioural changes, the social effects and contribution to the knowledge base. Aspects of many policies may take considerable time to become visible, are highly dependent on the specific social and economic context and the complexity of attribution to the scheme increases with time. The conclusion is, however, that there is no one right method for impact assessment and that diversity in methods is an intrinsic characteristic of good evaluation practice.

7.1.8 Contributions and applications of qualitative methods

If qualitative methods are essentially a means of finding out what people do, think and know, it follows that they may not be the primary method for all evaluations. However, as Scriven (1991) argued, a substantial part of a good evaluation is wholly or chiefly qualitative since regardless of the method of data collection used there must always be description and interpretation.[12] Qualitative

[11] See section 3.2.2.4 for a discussion of process in qualitative research; [12] See sections 3.1.1 and 3.2.2.2.

approaches are particularly important in the preliminary phases of research,[13] and in some instances a qualitative evaluation will be a precursor to a quantitative study. Because qualitative methods are both understandable and accessible they are also valuable in participatory/collaborative evaluations.[14] They are likely to be helpful where the issues are complicated, numbers small or research subjects unusual.[15]

The following areas are the ones in which they are considered to have particular strengths.

Discovery: It is likely that a social agenda will almost never exist as a choice between fixed alternatives – there is a continual search for alternatives more acceptable than those proposed at the outset. A study limited to the original narrow questions ignores much that may be important. As Cronbach and co-workers (1980) argued, the community is at least as well served by the discovery of new possibilities for action as by the definite appraisal of a programme fixed upon in the past. The use of qualitative research methods should aid this discovery. They may also be useful in the identification of natural solutions to problems – those that people devise for themselves without policy intervention (Marshall and Rossman, 1989).

Qualitative methods are also appropriate where project outcomes are uncertain or not carefully articulated.

Process studies and evaluations focus on how something happens rather than the results obtained. Qualitative inquiry is particularly suited to this as a characterisation of process requires detailed familiarity with all aspects of the programme, in particular a sensitivity to unanticipated events.[16] It also acknowledges the multiple experiences of participants and in many evaluations their perceptions are likely to be a main concern. Process evaluations illuminate how a programme works and aim at understanding its internal organisation – in this way are they are useful for stakeholders not closely acquainted with its detailed operation. Process data should allow fairly early judgements to be made about whether the programme is working in the way intended. **Formative evaluations**, intended to improve a programme, particularly rely on process data.

The linkages between processes and outcomes are fundamental in many evaluations. In many evaluations there will be a clear emphasis on outcome. However, even where outcomes are clearly specified and can be assigned one or more measurable values it is still necessary to understand why and how these are achieved. That is, there is still a need for accompanying "well-planned and conscientiously executed" process evaluation, which also acts as "a critique of the measurement process and the experimental arrangements" (Campbell 1978:196). Process evaluations may also be required to facilitate replication of a successful programme at some other site. Where outcomes are ambiguous, it is similarly important to have some explanations as to why this is the case.

Individual outcomes: Patton (1990) also urged the use of qualitative methods in the evaluation of individual outcomes. Much current service delivery is couched in terms of meeting individual client needs. A thorough evaluation should provide some demonstration of these qualitatively different outcomes. This may resemble a **case study** evaluation, though this term is also applicable when the case is a programme, an organisation or a community.

Comparative analysis: Keen and Packwood (1995) focused on the case study evaluation of implementation. They pointed out how studies are usually designed to incorporate some kind of comparisons, for example, between different approaches to implementation or between intervention and non-intervention sites. The examples provided are of studies of the implementation of GP fundholding and of resource management in six pilot hospitals. Both used a variety of methods, a distinctive though not a unique feature of case study research.

Implementation: The diversity – or the comparisons – central to the case study approach outlined above is another theme picked up by Patton who commented on the importance of adapting basic programmes to local community needs and circumstances, an issue equally applicable to the UK. To understand this a holistic evaluation picture of each site is needed. Central to this is **implementation evaluation**. To know if a programme is effective after it is fully implemented is important, but it is also necessary to know the extent to which implementation has been achieved. Patton (1986), in his advocacy of useful evaluations, even suggested that if resources were limited and a choice had to be made, implementation information is more

[13] See section 3.2.2.5; [14] See section 3.1.3; [15] See section 3.1.1; [16] See section 3.2.2.4.

valuable than outcome information. Unless the programme is found to be feasible and working as intended there may be little reason to look for outcomes. Implementation evaluation should include data relating to structure, inputs and process.[17]

Steckler (1989) described the study of implementation as that of testing internal validity (defined as the degree to which the results obtained from an evaluation can be attributed to the intervention). Referring specifically to health promotion programmes, but equally applicable to other types, he asserted how rarely these can be dichotomised into existing or not existing – rather they are characterised by the degree to which they have been adequately implemented. Health promotion research needs equally to address questions relating to why effects occurred (or not), as well as to whether (or not) they did.

7.1.9 Receptiveness to qualitative evaluation

Some areas of study have been more receptive to the use of qualitative methods than others. As noted by Holland (1983), much health promotion/education research depends on such approaches. Shiroyama and co-workers (1995) have written about recent experiences of evaluating Scottish health-promotion projects taking place in primary care. They suggested that the "multi-faceted nature of many health promotion interventions and the emphasis on qualitative behavioural knowledge based on attitudinal changes make the application of experimental techniques difficult". But, at the same time, short-term funding of projects has encouraged the use of pre-test–post-test designs rather than detailed, and potentially more useful, process measures and accounts. They tabulate useful, and comprehensive – in that they refer to factors in addition to methodological choice – recommendations for health promotion evaluation in primary healthcare settings. Many of these are likely to be applicable to other evaluations. In effect this follows Scriven's (1991) standards for a good evaluation. At the design stage of an evaluation evaluators should consider:

- **the setting**: staff, culture, communications, history, relevant legislation
- **roles and relationships**: including those between evaluators and project team
- **aims and objectives**: clarity, value of qualitative objectives and outcomes
- **research methods**: suitability, potential for adaptation
- **ethical concerns**
- **costings**: extent to which cost-effectiveness is able to be determined
- **feedback and dissemination**

(based on Shiroyama *et al,* 1995:232).

Qualitative research strategies have also been applied to the field of community development for health (Beattie, 1995). Such methods are said to be useful because they offer a way of conducting evaluations which reflect this type of initiative: an emphasis on process, working in non judgemental ways, sensitivity to local cultures, negotiating concerns with participants, shared ownership of data and planning for the future. In addition, Beattie listed evaluation styles prominent in community development for health projects in the UK and provided useful references.

Good (1992) discussed the adoption of qualitative methods in assessing interventions in North American family medicine, resulting from the realisation that quantitative models are limited and lack meaning. They "fail to tap into the complexities of medicine's clinical tasks and challenges". She extended this to the experiences of physician researchers in Third World countries who, finding that the problems there defy assessment by conventional epidemiological means, are increasingly influenced by medical anthropology and qualitative methods of study. Bennett (1995) gave a number of examples of international organisations encouraging the use of qualitative methods in research and service activities. As a counterpoint to this, Barker (1995) examined the constraints on their use in the developing world. She argued that the research process itself militates against the various qualitative research approaches which would frequently provide the appropriate tools for the study of management problems. More particularly, much of the research is evaluative in nature and formal evaluation with its predisposition to concentrate on the achievement of formal goals, tends to rule out finding out what is really happening, or the learning necessary to effect improvement.

Reinharz (1994) has written on qualitative evaluation and policy with particular reference to research on the elderly, illustrating its specific strengths in this context. Some of the studies she reported were not commissioned as evaluations

[17] See also section 3.2.2.4.

but act as evaluations of policies in, for instance, the application of different levels of medical care for Alzheimer's patients. One case study became a qualitative evaluation of an intensive home care support programme. More generally the strengths are seen as the provision of individual accounts, which simultaneously enable the voice of the underprivileged to be heard, and the avoidance of ethical problems which may follow from the imposition of an experimentally-designed evaluation.

7.1.10 Problems of qualitative evaluation

Many would regard qualitative evaluation as a necessary component of any effort to reach and understand decisions in health programmes. It aids understanding of the complex contexts of patient care, health education and programme administration and investigates the human actions and decisions made as part of these programmes (Pearsol, 1985). Such evaluation is not without problems.

Stakeholders. Approaches relying on stakeholder accounts assume that these individuals (or groups) can readily be identified, recognise their need for the information to be produced by the evaluation, and share ownership of the programme and its goals. All of these assumptions are questionable (Weiss, 1984). Stakeholders are not always identifiable at the start of an evaluation. Not all want the programme to be evaluated and/or are not willing to participate in the process. The results may be seen as a threat to organisational interests rather than as advice or guidance. Neither do stakeholders necessarily know in advance what they will need to know in order to make a decision. And, as Weiss pointed out, not uncommonly people do not actually want to know anything – but neither do they want to admit this. Stakeholders are not uniformly influential and some will have no decisions to make as the result of the evaluation. However, Weiss concluded that the stakeholder approach seems to hold "modest promise for achieving modest aims". It can improve the fairness of the evaluation process but a stakeholder evaluation will not bring "harmony to contentious program arenas". Diverse views may be elicited but not necessarily contained.

Time and resources. On a practical level a qualitative evaluation is time consuming to produce – evaluators need enough time to undertake the field work and probably an equivalent amount of time to synthesise their findings into a final report. In particular, stakeholder approaches both impose an increased burden on the evaluator and demand time and attention from others (Weiss, 1984). Data collection costs per respondent are likely to be far greater than if a survey questionnaire were to be used (Reichardt and Cook, 1979). The report may also be lengthy and not readily reducible to a short executive summary. For some readers the style adopted may be unfamiliar.

Generalisability. It is likely that the main criticism of qualitative evaluations will be related to the lack of generalisability.[18] As Barker (1995) has persuasively argued, evaluations seek to produce knowledge specific to people and to programmes. It could also be argued that generalisations are not necessary to analyse what is going on in a project or programme at any particular time and to decide what actions must be taken (Barker, 1995). Scriven (1991:209) regards generalisability as the equivalent of **potential** or **versatility** and versions of a programme, or parts of it, may well be able to be adopted elsewhere. Evaluations are often conducted on trial programmes operating with highly committed staff members and considerable management support – it is particularly important to estimate how they would work under less than optimum conditions. Stake (1986:x) also wrote of the **reader-made generalisations** that can be made as a result of providing rich and detailed accounts.

> I have tried to emphasise the uniqueness of this case more than its generality. I have paid less attention to what ... is common to other evaluation content, more to its special context. Believing that each reader will generalize to sites and circumstances about which I know little, I have tried to provide great detail about particulars that facilitate those reader-made generalizations.

In other words the reader, not the researcher provides reference population and reference groups.[19] The evaluator is responsible for writing in a maximally comprehensible way and for the interpretation, which assists the reader's intuitive analysis and generalisation (Shadish *et al,* 1991:271).

This is rather different from the obvious desire of some commissioners for 'ready-made' or 'toolkit' type solutions, where a report describes a programme that can simply be imported or bolted on to an existing organisation. However, it may be much more realistic to describe the relationship between a programme and its context, in the

[18] See sections 4.1 and 5.4; [19] See section 4.1.1.1 for a critique of this position.

knowledge that every context is somewhat different – different resources, different micro-structures, different actors, etc. – and that a programme is unlikely to work in exactly the same way with exactly the same effect unless some attention is paid to its adaptation to the new environment. In this sense, qualitative evaluations can be seen as supporting the element of discretion and judgement, which is critical to successful management, rather than as technologising or automating it.

7.2 Methodological analysis of a sample of evaluations

The purpose of this section is to examine some recently published evaluations in the field of health care, focusing particularly on methodological issues. That is:

- to describe some of the ways in which qualitative methods are currently being used, either alone or in combination with quantitative methods
- to compare and contrast the respective uses of qualitative and quantitative methods, assessing the strengths and weaknesses of these approaches
- to identify any potential uses for qualitative methods.

Searching research literature by reference to methods, rather than by reference to a substantive topic area, is, in general, an impossible task: databases rarely use methods as key terms; methods are often inadequately described; and methods are not always specifically acknowledged.[20]

The approach used in this case study was to carry out a hand search of the issues published in 1995 in a range of journals in which one might expect to find health programme evaluations. The selected journals were:

British Journal of General Practice
British Medical Journal
Health Education Journal
International Journal of Nursing Studies
Journal of Advanced Nursing
Journal of Public Health Medicine
Social Science and Medicine.

(All references in the rest of this chapter are 1995 publications unless otherwise stated.)

Patton (1986:31) reported adopting two simple criteria in identifying a sample of health evaluation studies:

- an actual operating programme existed
- systematic data had been collected.

Fewer than half of his original list of evaluations met these standards. Applying these criteria to the present case it is likely that few studies would fit. Instead, the strategy used was to locate all research reports that were explicitly described as evaluations, but to include also examples from other studies which, given the range of definitions and types outlined in the first section, could quite reasonably have been characterised in this way. The emphasis has been on studies that have either evaluated some organisational change or have introduced relevant methodological issues. Consequently, a number of studies defined in their titles as evaluations, have been omitted from the sample. Twenty-six studies were finally selected for inclusion. Detailed summaries of these studies are given in *Table 3*.

Some of these papers (Jones and Mullee; van der Walt *et al;* van Teijlingen *et al)* were written primarily to address the methodological problems of the evaluations to which they referred. A significant number of other articles – conference proceedings, discussion papers, editorial comments and correspondence – focused on, or made reference to, relevant methodological concerns. Although they have not been directly used in this analysis this would seem to be one of the incidental benefits of using the hand search approach.

7.2.1 Classification of evaluations

7.2.1.1 Exclusively quantitative

- Two of these evaluations were conducted as RCTs (Frost *et al;* Lindholm *et al*). In both cases the interventions were relatively discrete: a fitness programme for patients with low back pain and a programme of intensive healthcare advice given to subjects with multiple risk factors for cardiovascular disease. The first evaluation assessed patients using validated measures before and after treatment, pain diaries (scaled) and a 6-month post-treatment questionnaire. The second used clinical measurements of blood cholesterol together with patient completed questionnaires relating to life style and diet.

[20] The only exception might be where a methodological innovation is the topic of a paper intended to describe and promote it. Once it becomes routine practice, a method becomes much less visible.

TABLE 3 *Summaries of evaluation studies published in 1995 in seven selected journals*

Study	Purpose of study	Study design, data collection method, sample type, sample size, response rate, date of data collection	Notes/comments
Blakey and Frankland *Health Educ J* 1995;**54**:131–42	Evaluation by an outreach worker of HIV prevention for women prostitutes	Preliminary research – contacts with other projects. Data collected was quantitative and qualitative. Contact sheets completed for each contact. Significant discussions recorded (134 with 87 different women). Outreach worker completed weekly diary – to record feelings and events. Structured interviews carried out with 50 of the contact population (non-representative sample, but quota and time samples) on relevant health issues and views of project. 1990–93	Ample contextual detail provided. Results included tabulation of contacts and diary quotations. Discussion of the lessons learned from the project including need for careful planning and value of visits to other schemes. Action research focusing on prostitution likely to meet difficulties that do not arise in most research settings.
Bruce and Griffioen *Soc Sci Med* 1995; **40**(8):1109–16	Evaluation of service developments in infant feeding – appointment of baby feeding adviser and other policy changes	Pre- and post-intervention studies (1988 and 1990). Two surveys at both points: non-medical staff and mothers from maternity unit of DGH. Staff: self-completion questionnaires, 48 (80%) returns (1988), 65 (87%) returns (1990). Additional comments incorporated in analysis. Mothers: interviews (250) and 6-weeks post-natal mailed questionnaire. 1988	Detailed paper in which non-experimental design justified. Reasons for not using RCT explained, e.g. resources, inability to influence developments in control hospital. Some contextual detail given but limited detail about process, e.g. work of baby feeding adviser. Said to provide valuable information for hospital staff to use in development of service changes.
Chrystie *et al.* *BMJ* 1995;**311**:928–31	Pilot study of voluntary, named testing for HIV in a community-based ante-natal clinic	Pre-study interviews with clinic staff. 126 women attending one group practice offered test. 44% uptake. Midwives completed questionnaires relating to time spent and other problems with booking process. 1993–94	Main problem said to be change in dynamics of booking visit and in midwife–client relationship. Does not seem that all the evidence for this would come from questionnaires and no indication how views of women elicited.
Denman *et al.* *Health Educ J* 1995; **54**:3–17	Measuring the impact of a theatre in HIV/AIDS education programme on the knowledge and attitudes of Nottinghamshire secondary school students	Used combination of qualitative and knowledge and attitude to HIV and AIDS of 13/14-year-olds assessed before and after intervention using confidential, self-completed questionnaire. 252 in experimental and 428 in control group did both tests. Additional questions re-acceptability asked of experimental group. Full programme observed to monitor issues raised and to assess interest; focus-group interviews to examine in detail attitudes to HIV/AIDS. Teacher survey using self-completion questionnaire.	This paper focuses on the first part of the research – pre- and post-test data. The setting said to place limitations on the evaluation design, i.e. the inability to conduct a completely randomised trial. Knowledge found to increase in several areas and attitudes also influenced. Children were receptive to the programme – comments listed. Researchers' interpretations presumably influenced by content of focus groups, though this not reported here.
Dujardin *et al.* *Soc Sci Med* 1995; **40**(4):529–35	An evaluation of women's compliance with recommendations to give birth in hospital, following the identification of various risk factors	Ante-natal records of 5060 pregnancies from 11/20 health centres in Zaire examined. 1988	Referral success rate was only 33%. Seven hypotheses tested; most important were woman's perception of risk and geographical accessibility of hospital. Said that research is too general to clarify the real explanations and that additional qualitative work (open interviews, focus groups) should be the next step in analysing the problem.
Frost *et al.* *BMJ* 1995;**310**:151–4	Evaluation of a fitness programme for patients with low back pain	Single, blind RCT involving 81 patients recruited between 1991 and 1993. Assessments carried out before and after treatment using a number of validated measures. Pain diaries kept with numerical scale between 0 and 100. Postal questionnaire sent 6 months after 2nd assessment	Differences between the two groups but validity of long-term follow-up questioned because a number of patients crossed over from the control group. Suggested that such a change in treatment indicative of superiority of on treatment group over another and could be used as an outcome measure. Informal comments of patients at 6-month assessment reported. Authors propose further trials in other centres.
Gillam *et al.* *Br J Gen Pract* 1995; **45**:649–52	Evaluation of an outreach model of ophthalmic care in general practice in terms of impact on GPs, use of secondary ophthalmic services, patient views and costs (pilot study)	Prospective study of 17 practices in London. 17 matched-control practices used in comparison of referral rates. Activity data collected in outreach clinics. Questionnaire sent to 55 study GPs (85% response), semi-structured interviews with one GP from each study practice (asked for views on suggestions for development). Self-completion questionnaires given to 210 patients from six study practices (75% response) and 246 patients at hospital ophthalmology outpatient department. (61% response). 1992–93	Educational impact on GPs not impressive. Unit costs three times more than conventional outpatient treatment (low patient throughput). Referral rate to hospital lower for study practices. Journey and waiting times shorter for patients attending outreach clinics. Majority satisfied with service. Service continued with fewer satellite clinics serving more practices. Little detail about qualitative aspects of evaluation. Overall a comprehensive evaluation but benefits of scheme limited. Efficiency data probably had most impact.
Hudelson *et al.* *Soc Sci Med* 1995; **41**(12):1677–83	Study of how families perceive and respond to children's acute respiratory infections in order to develop relevant health programme	Two rural communities, linguistically and culturally representative of larger population block, served by public clinic where acute respiratory infections programme planned and where acute respiratory infections mortality and morbidity high. Described as a focused ethnography study – used semi-structured interviews with key informants; narratives of past episodes; clinic interviews with mothers of current patients; hypothetical illness scenarios and interviews with health practitioners. 1991	Study design highly structured but researchers encouraged to adapt this to suit local conditions. Ethnographic techniques used to address issues of relevance by providing in-depth local perspective.

continued

TABLE 3 contd *Summaries of evaluation studies published in 1995 in seven selected journals*

Study	Purpose of study	Study design, data collection method, sample type, sample size, response rate, date of data collection	Notes/comments
Jones and Mullee *Br J Gen Pract* 1995; **45**:497–9	Study of nurse-run asthma care in general practice	A prospective study focusing on outcomes of asthma care in two general practices. One with nurse-run care, second was well-matched control with traditional (reactive) approach. 100 patients (5–65 years) in each. Data collected from interviewer-administered questionnaire, lung function measurements and extracts from case notes collected in three phases. Pre-12-month facilitation period in intervention practice, at end of this, and 12 months later. Practice team interviewed confidentially in second phase. 1988–91	Although considerable resources invested in the study unable to demonstrate statistically significant differences. Observational work, knowledge of working arrangements and staff interviews indicated some of the likely reasons. Also suggested that the control practice too advanced in the way in which it managed asthma. Authors conclude that may be impossible to conduct RCTs in this field. Propose that a more useful approach would be to use qualitative interviews with key informants and/or patient focus groups to evaluate patients' perceptions of proactive care.
Leese and Bosanquet *BMJ* 1995;**310**:705–8	An evaluation of the changes that have taken place in general practice structure and organisation between 1986 and 1992	26 group practices in one family health service authority in 1992 compared with 1986 study of same area when practices designated as innovators, intermediates and traditionalists. One partner from each practice took part in interview and structured questionnaire completed relating to structural and service features of the practice. GPs views concerning 1990 contract recorded (five-point scale)	All practices had invested heavily in equipment and services but differences remained. These depended on geographical location and practice philosophy. Results said to suggest some conflict between professional identity and public interest. If the latter factors were to be explored further would need to use qualitative methods.
Lindholm *et al.* *BMJ* 1995;**310**:1105–9	Evaluation of the additional benefit of intensive healthcare advice given during six group sessions to subjects with multiple risk factors for cardio-vascular disease	Prospective, RCT. 681 subjects aged between 30 and 59 years from 32 Swedish health centres. Percentage reduction in cholesterol concentration measured. Questionnaires relating to life style and diet completed by participants at the group sessions. 1990–93	Limited additional benefit gained from being in the group receiving the intensive healthcare advice. Authors concluded that better methods of communicating the messages needed to be devised and that their delivery should be customised. Following up these suggestions would imply the use of qualitative research methods.
Lord and Green *Health Educ J* 1995; **54**:453–64	Evaluation of an 'exercise on prescription' scheme	Series of self-completion questionnaires including validated instruments for participants. Administered at initial referral, (252/419 patients attending) 10 weeks (n = 77) and 6 months (n = 64). Consultation sheets maintained by health and fitness workers. GP referral records analysed. Additional qualitative work with participants (focus groups, including one for non-attenders), GPs (semi-structured interviews with the practice) and steering group members (focus group). Financial records kept. 1992	Pluralistic evaluation incorporating different perspectives of key stakeholders. Primary compliance (attendance) with scheme thought reasonable but relapse a matter of concern. Possible explanations emerged from qualitative data – with useful practical suggestions for improvement. Data analysis from questionnaires mentioned but not that from qualitative work. (Qualitative) methodological issues treated more superficially. Nevertheless results appear to have been influential.
Mathews *et al.* *Soc Sci Med* 1995; **41**(12):1715–24	A formative evaluation of an AIDS education programme for secondary school students in South Africa	(Quantitative survey of students' attitudes, knowledge and behaviour not reported in this paper.) Focus groups with students, focus groups and free-attitude interviews with teachers in one secondary school. Tapes transcribed and analysed. 1991–92 (Final phase, piloting lessons, evaluation by interview and questionnaire not reported)	Teachers did not implement programme as intended. Researchers thought they had understood and acknowledged teachers' concerns but had not recognised importance of religious values. Survival of project depended on further qualitative research. Aim had been to develop national resource and programme now being assessed in another setting, involving teachers as co-researchers.
McKenna *et al.* *Int J Nurs Stud* 1995a; **32**(1):79-94 *Int J Nurs Stud* 1995b; **32**(1):95–113	Evaluation of the implementation of a nursing care model for long-stay psychiatric patients First paper covers literature review and methodology; the second presents and discusses findings	Random selection of hospital and ward for study. Matched control ward from same hospital. Quasi-experimental design adopted using action research approach involving patients and staff, the latter providing change agents as study progressed. Various instruments used before and after (× 2) implementation to measure dependent variables. Patient records analysed. Pilot study carried out on adjacent ward. Pre-intervention observations carried out to understand routines and to lessen any reactive effects. 1989–90	On experimental ward some statistically significant improvements. Alternative explanations for these significant findings discussed and mainly considered to be implausible. Non-significant differences between study and control wards similarly discussed. Limitations include lack of generalisability from study – possibly from any action research project. Authors acknowledge lack of open-ended questions and/or use of ethnographic methods, which could have provided richer data.
Newens and McEwan *J Adv Nurs* 1995; **22**:267–75	An evaluation of HIV and AIDS awareness sex education programmes for young people with severe learning difficulties	Non-participant observation of five randomly selected sex education sessions in each of two special schools. Field notes and tape recordings (transcribed). No date given	Setting and content of courses differed. Described and analysed in detail and relative merits discussed. Presented within a framework of 'rights' for those with learning difficulties.
Paterson and Peacock *Br J Gen Pract* 1995; **45**:255–8	A description and evaluation of the integration of complementary practitioners and therapies into a general practice	Used qualitative methods influenced by a model of co-operative inquiry. Practice team defined what they would like from study. Different interview schedules for the different groups in practice. Responses transcribed and main issues identified and described. (Sept 1993–Jan 1994) 12 months (1992–93) quantitative data for patients attending complementary practitioners assembled. Practice organisation described by researcher	Claimed that this method of inquiry has promoted greater discussion, understanding and occasionally problem resolution. Authors say that descriptive and qualitative research on a single practice likely to be affected by desire for success. To counteract this used external researcher and included multiple perspectives, though acknowledge lack of patient participation in the study.

continued

TABLE 3 contd *Summaries of evaluation studies published in 1995 in seven selected journals*

Study	Purpose of study	Study design, data collection method, sample type, sample size, response rate, date of data collection	Notes/comments
Prasad and Costello *BMJ* 1995;**310**:621–3	Evaluation of the impact and sustainability of a 10-day training programme (for doctors, midwives and nurses) on the benefits of early breast feeding	Intervention study. Infant feeding habits assessed by maternal interview. Three groups of mothers included: control (n = 172), immediately following intervention (n = 195) and 6 months later (n = 101). 1992–93	Observations carried out in hospital. Interviews used a short semi-structured questionnaire. Data collected were mainly quantitative and tabulated. (Many subjects were illiterate). Early post-intervention group started to breast feed earlier and less likely to use pre-lacteal feeds but impact fell off among late follow-up group. Focus group discussion with staff conducted to investigate difficulties in sustaining health education programme.
Smith *et al.* *BMJ* 1995;**310**:1175–78	Evaluation of a (pilot) therapy service for adults who had experienced sexual abuse as children	Questionnaire survey. Three standardised psychological questionnaire administered at beginning and end of study. Initial questionnaire about previous health service use incorporating some free text responses. Five-point satisfaction scale with service received. 92 clients took part. 59 completed final tests. Sept 1993–August 1994 Medical records of 18 clients compared with those of 36 age- and sex-matched controls to determine differences in health service use. 1993–94	Clients may not be representative but said to be impractical and unethical to randomly distribute clients into treatment and non-treatment groups. Improvements shown on psychological scores and clients satisfied with service. Clients heavier users of other health services with more difficult-to-manage conditions. Described as the first evaluation of such clients previous use of resources.
Spiers and Jewell *Br J Gen Pract* 1995; **45**:31–3	Evaluation of the use of a generic counselling provided by a single counsellor working in two general practices	External evaluator worked with family health serivce authority and GPs before counsellor appointed to clarify aims of service and methods of study. 293 patient studied. Patient data collected over 2 years. Indications for referral, counsellor's initial assessment, immediate post-treatment outcome (counsellor assessment) plus outcome 1 year later from review of medical records. Type/dosage of any psycho-tropic prescribing recorded. Patients attending in first (pilot) year sent questionnaires 6 weeks after final counselling session (84% response rate). Questionnaire completed by non-medical staff in practices (100% response). GPs interviewed, and also provided data on proposed alternative treatment in absence of counselling service. 1990–92(?)	Authors conclude that the study shows the value of clarifying referral criteria and intended role of counsellor before introduction into general practice. Little detail of questionnaires used – for patients relate to helpfulness of service and willingness to use again. Other patient outcomes defined by professionals not patients. Non-medical staff asked if they knew what the counsellor did and if service should continue. No details of GP interviews other than that attachment considered very successful. Data said to be descriptive but findings robust.
Stewart-Brown *et al.* *BMJ* 1995;**311**:1543–7	A study to investigate the effects of general practice fundholding on prescribing habits	Prospective observational study of 12 practices in Oxford region (five non-dispensing fundholding, three dispensing fundholding, five non-fundholding). Prescribing analysed (costs, numbers, generic usage) for same 6-month periods of 3 financial years (1990/1, 1991/2, 1993/4)	Concluded that initial changes in prescribing habits attributed to fundholding not sustained over the longer period. Authors point out that number of practices small and results may not be generalisable to all types of practice. Two of non-fundholding practices preparing for fundholding by 3rd year. Differences between dispensing and non-dispensing fundholders observed. One explanation proposed but no data to support this. Potential influences on prescribing behaviour detailed but this study did not set out to explore these further. Demonstrates value of longer term studies.
Tudor-Smith *et al.* *Health Educ J* 1995; **54**:393–404	An evaluation of a 'decisions' programme' on substance use provided by a mobile health education resource for children	Non-experimental pre-and post-intervention design, collecting complete data from 339 10/11 year olds in Wales. Self-administered questionnaire completed on three occasions – before intervention (60–90 minute lesson), up to one week after and 2 months later. Knowledge, beliefs, skills, intentions and behaviour included, with post-test questions asked about the programme. Results suggested acceptability of programme and increased knowledge. No change in intention to smoke or to choose other positive actions for health. 1993	Limitations ascribed to non-experimental design. Unable to attribute with certainty observed changes to the intervention. Group (67%) completing all three surveys also thought to have different characteristics from those pupils who did not. Within this design no possibility of following up some of the factors of concern, e.g. relation to positive images of smoking and drinking, or to ask for suggestions for improvement in the lesson. Authors acknowledge difficulty of isolating specific contributions of programme to national health goals.
Twinn and Shiu *Int J Nurs Stud* 1995; **33**(4):442–54	An evaluation of the effectiveness of public health nursing through a study of maternal and child health centres in Hong Kong	Multiple-case study design (four), each focusing on nursing staff and clients, using qualitative and quantitative methods. Staff: questionnaires for all (n = 42), from these purposive sample (n = 16) interviewed (semi-structured, taped). Clients: telephone interview with systematic sample of 21 (64%) parents of non-attenders, face to face semi-structure taped interviews with sample of attenders using set criteria for selection (32 parents). Epidemiological survey to determine health needs of women and children in catchment area	An evaluation of quality of care rather than a programme evaluation but interesting use of case study analysis. Problems with defining outcome in long-term, preventive and community services discussed. Issues of reliability and validity considered. Use of multiple methods said to provide opportunity to evaluate consistency of data from staff sample. (Methodological paper published elsewhere.)

continued

TABLE 3 contd *Summaries of evaluation studies published in 1995 in seven selected journals*

Study	Purpose of study	Study design, data collection method, sample type, sample size, response rate, date of data collection	Notes/comments
van Teijlingen *et al* *Health Educ J* 1995; **54**:357–66	Discussion of methodological and practical problems that occurred during the evaluation of an anti-smoking campaign for school-age children	A longitudinal study (1988, 1989/90, 1991) of members of Smokebusters Club. Structured self administered questionnaire completed in classroom in absence of teacher. 10% sample of three school year groups, primary class from structured sample of schools, secondary pupils from random sample of pupils in all schools	Face-to-face interviews planned but not completed due to illness. Intention to compare national data on smoking prevalence but not available for relevant year. Peer group influence not evaluated. Meaning of being a club 'member' not explored. Longitudinal survey meant questionnaires had to be named. Affects likelihood of truthful replies. No social class or educational data permitted. Mentioned, but did not establish, importance of other initiatives. Too many variables to be allow conclusions about cause and effect. (Evaluation published elsewhere.)
van der Walt and Mathews *Soc Sci Med* 1995; **41**:1725–9	Exploration of response of health services managers to an earlier qualitative evaluation of a community health worker programme in South Africa	Process evaluation carried out using observation of home visits, free-attitude interviews with random selection of community members visited, focus-group discussions with health workers. 1992. Managers interviewed after report issued and some months later. (Recorded, transcribed, analysed according to themes)	Original research report made recommendations in 12 areas. None reflected in subsequent actions. Main reason said to be pressure from influential local sources. These had not been included in original evaluation, which had interviewed 'users' and not political actors. Said that qualitative researchers need to research broadly as well as 'in-depth'.
Wellard *et al.* *J Adv Nurs* 1995; **21**:737–42	Study of the introduction of evaluation of clinical educators in a school of nursing	A responsive design. Semester 1: Individual co-structured interviews (discussions) with stakeholders. Revision of position statement. Semester 2: Preparatory training. Agreement on two evaluation methods – self evaluation of role performance and observation of teaching by evaluator	Design clearer than the precise methods used. Evaluation process, which involved all stakeholders assisted in identifying areas to be developed. Collaboration said to be beneficial to all.
Wilton *BMJ* 1995;**310**: 369–72	Study aimed at obtaining the views of house officers who had rotated through general practice in their pre-registration year. (Additional comparison of type and hours of pre-registration work in general practice and hospital)	Postal questionnaire sent to 28 pre-registration house officers in general practice (1981–91). Structured and free-text questions to obtain views about the rotation. 93% response rate. (Self recording of working hours and duties four based in general practice, eight in hospital)	GP rotation scheme had been evaluated previously (1985). This work supplemental and able to incorporate subsequent career experiences of participants. Second part of study used as additional evidence to support value of scheme in terms of clinical content and specific teaching.

- Other studies, although collecting only quantitative data have used a variety of non-RCT designs, mainly some variation of pre- and post-intervention testing, with control data for most (McKenna *et al*) or some part of the study (Smith *et al*; van Teijlingen *et al*).[21] Some studies have not used any control data (Bruce and Griffioen; Tudor-Smith *et al*). Bruce and Griffioen actually described their research design as **non-experimental**: in this case using pre- and post-intervention staff and maternal questionnaires after changing infant feeding policies in a maternity unit. Together with the RCTs this group of evaluations depend on some form of statistical analysis.

  Although classified here as an exclusively quantitative study, McKenna and colleagues preceded their more experimental work – the introduction of a nursing care model for long term psychiatric patients – with a period of observation. This was intended to minimise subsequent reactive effects rather than constitute a method of research.
- A number of studies have relied on the analysis of existing, or routinely produced, data and have not been designed to generate any other material. For example, Stewart-Brown and colleagues, in their investigation of general practice prescribing habits following the introduction of fundholding, conducted a prospective observational study of eight fundholding and five non-fundholding practices, using the Prescribing Analysis and Cost System data.[22]
  The research reported by Dujardin and co-workers. relied on a retrospective examination

[21] Categorised here as a quantitative evaluation, the study design planned by van Teijlingen and colleagues did originally include some qualitative work but this aspect was never implemented; [22] Although the term 'observational' may be used to refer to non-RCT quantitative, epidemiological methods it has not been used in this classificatory scheme because of potential confusion with observation as a qualitative method of research.

of 5060 ante-natal records to assess the extent to which recommendations about preferred place of delivery were followed (Dujardin *et al,* 1995).
- Structured questionnaires were used in Wilton's study of pre-registration house officers in general practice placements. He used a postal questionnaire for this group, but, collected additional data relating to working activities/hours. The second part of the study used similar material recorded by hospital based house officers as control data. Leese and Bosanquet undertook a comparative study of a number of general practices before and after the introduction of the 1990 contract, using a interviewer administered questionnaire. Although in both of these studies many of the results were presented in numerical form, there was no statistical analysis and the major discussion elements relate to the policy issues.

7.2.1.2 Mixed method studies

Ten of the studies examined used both quantitative and qualitative methods, though it seems relatively unusual for both to be given equal attention – at least in the published papers. One exception might be the multiple case study design used by Twinn and Shiu in their evaluation of the effectiveness of Hong Kong maternal and child health centres.

Those evaluations in which qualitative approaches seem to dominate are typically service reorganisations where considerable emphasis is placed on interview and observational data and quantitative information is extracted from relatively simple recording systems (e.g. Paterson and Peacock; Spiers and Jewell). The study by Gillam and co-workers of outreach ophthalmic clinics in 17 study (and 17 control) general practices collected a greater range of numerical activity data. Chrystie and colleagues reported a pilot study of the introduction of voluntary named HIV testing in a community ante-natal clinic with data derived from time measurements, interviews and questionnaires. Although neither of the latter methods is defined in detail the policy concerns are fully explored. Blakey and colleagues used worker diaries, as well as records of contacts and structured questionnaires, in their evaluation of an HIV/AIDS prevention programme. Lord and Green's evaluation of an exercise on prescription scheme made use of validated questionnaires, consultation and referral records in addition to focus groups and individual interviews. Other studies have used interview and observational data together with clinical measurements – for example, a study of nurse-run asthma care in general practice, which particularly focused on measurable outcomes (Jones and Mullee).

Mixed method evaluations in which quantitative data predominate tend to be those which employ an experimental pre/post-test design with some type of qualitative work as an adjunct. For example, Denman and co-workers studied the impact of an HIV/AIDS theatre-in-education programme, assessing knowledge and attitudes before and after the performance. They also conducted focus group discussions with pupils, though this aspect of their research is not reported in the article. In contrast, the paper by Mathews and co-workers similarly dealing with AIDS prevention education and using surveys to ascertain students' attitudes, knowledge and behaviour, addresses only the qualitative aspects of their study.

7.2.1.3 Qualitative studies

Fewer evaluations/studies used solely qualitative methods. The research by Hudelson and co-workers is described as **focused ethnography**: families in rural communities in Bolivia were studied in order to discover the ways in which they perceived and responded to children's acute respiratory infections. Research methods were diverse – semi-structured interviews with key informants, clinic interviews with mothers, interviews with health practitioners, studying narratives of past episodes of respiratory infections and using hypothetical illness scenarios. A study of two sex education programmes for young people with severe learning difficulties conducted by Newens and McEwan relied solely on observational methods, though this included audio-recordings of the teaching sessions.

Prasad and co-workers used observation, semi-structured interviews with mothers, and focus group discussions with hospital staff in their evaluation of a training programme in for doctors and midwives in Bihar, which aimed to promote the benefits of early breast feeding. However, despite the description of the interviews as 'semi-structured', they may only have been used in preference to a written questionnaire as there were doubts about literacy of the target population. The information obtained from the interviews appears in numerical terms, is tabulated and tested statistically.

A study, which could probably be regarded as the **evaluation** of an evaluation (van der Walt and Mathews), provides an account of a qualitative process evaluation of a community health worker programme which used free-attitude

interviews,[23] focus groups discussions and observations. They went on to explore the subsequent lack of response of health service managers to their recommendations. The main problem appeared to be that because the evaluation had focused on service users it had ignored locally important political actors.

Wellard and colleagues present a study of an evaluation system for clinical educators, showing how this was developed and introduced into a school of nursing. This used a responsive design incorporating co-structured interviews, observation and participant self-evaluation. *Table 4* summaries the qualitative methods used in these studies.

7.2.1.4 Quasi-qualitative data

In a number of studies free-text responses were used as part of a structured questionnaire (Denman *et al,* Smith *et al,* Wilton). These answers have been used to elicit the respondents' own views in their own words. In each case these responses have been given some prominence in the published report. For example, in the study of a pilot therapy service for adults with a history of child sexual abuse (Smith *et al*), some of these comments are tabulated to illustrate why previous services were though to be unsatisfactory. This is said to be the "first evaluation of such adults' previous use of resources".

TABLE 4 *Summary of qualitative methods used*

Qualitative method	Study (all 1995)	Study setting
Interviews		
Semi-structured	Gillam *et al*	GPs
	Lord and Green	GPs
	Hudelson *et al*	Key informants and health practitioners
	Prasad *et al*	Mothers of infants
	Twinn and Shiu	Clinic attenders and clinic staff
Free attitude	van der Walt and Mathews	Community members
	Mathews	Teachers in development of AIDS education programme
Confidential	Jones and Mullee	Practice team members
Cooperative inquiry	Paterson and Peacock	Practice staff
Co-structured	Wellard *et al*	Stakeholders
Not specified	Chrystie *et al*	Clinic staff
	Spiers and Jewell	GPs
Observation	Blakey *et al*	HIV prevention programmes
	Denman *et al*	HIV/AIDS education programme
	Newens and McEwan	HIV/AIDS awareness teaching sessions
	Prasad *et al*	Hospital routines
	van der Walt and Mathews	Home visits
	Wellard *et al*	Teaching sessions
Focus groups	Denman *et al*	Secondary school students
	Lord and Green	Attending participants, lapsed participants and project steering group
	Mathews *et al*	School staff and students
	Prasad *et al*	Hospital staff
	van der Walt and Mathews	Health workers
Diaries	Blakey *et al*	HIV prevention worker
Field notes	Blakey *et al*	Significant discussions recorded
Narrative analysis	Hudelson *et al*	Episodes of past respiratory illnesses
Vignettes	Hudelson *et al*	Hypothetical illness scenarios

[23] The free attitude interview technique requires the use of a single main question around which the interviewee is encouraged to freely explore his/her views and feelings. The role of the interviewer is to provide regular reflective summaries and to ask for clarification when required.

7.2.2 Assessment of strengths and weaknesses of qualitative and quantitative contributions

7.2.2.1 Control data

Many evaluations, not only the RCT studies have reported difficulties in maintaining the control groups, or in assuring their continuing distinctiveness. Frost and co-workers reported the crossing over of patients from the control group into experimental (exercise programme) group, even suggesting that in some circumstances such a change could be regarded as an outcome measure itself. Jones and Mullee, on finding few differences in patient outcomes in asthma care between a general practice where considerable effort had been made to implement a system of nurse-led care and a traditionally-organised practice, considered the possibility that the control practice was too advanced. Gaining cooperation from less well advanced practices is said, however, to be difficult. An investigation of prescribing habits following the adoption of general practice fundholding status, which used non-fundholding practices as controls, found that by the third year of the study some control practices were preparing for fundholding themselves (Stewart-Brown *et al*). Problems have also arisen when researchers have planned to use control material from other sources, for example, the Smokebusters Club evaluation described by van Teijlingen and co-workers intended to use national figures for smoking prevalence which were then not collected for the relevant year.

The evaluation of a mobile health education resource reported by Tudor Smith and colleagues did not, in fact, use a control group. However, there were indications those pupils in their sample who completed all three parts of their survey differed from those who did not.

7.2.2.2 Interpreting experimental studies

The RCT study on an intensive healthcare advice programme (Lindholm *et al*) did not report technical problems with the control group, but found that there were only limited advantages to the experimental group. This exemplifies the problem of studies that concentrate on input and output data only – if this is the case it is difficult to know how an intervention can be improved. Although this particular study apparently had no supporting qualitative components, the authors' own conclusions were that better methods of communicating relevant messages needed to be devised: it was necessary to learn about the best ways to influence people at risk and communications needed to be customised.

Other researchers have noted the lack of suitable explanations for some of their findings but do not specifically link this to their original methodological choices (Stewart-Brown *et al*). In this study of GP fundholding, one plausible hypothesis for the differential change in prescribing behaviours in dispensing and non-dispensing practices is offered: GPs in dispensing practices are likely to be better informed and thus enabled to respond more effectively to fundholding. However, it is likely that the use of some qualitative methods within this study could have led to alternate hypotheses being proposed.

The study of asthma care (Jones and Mullee) outlined above found few differences between study and control practices, though a long-term acquaintanceship with the planning and implementation process and the use of interview data meant that a number of likely explanations were forthcoming.

7.2.2.3 Inapplicability of the RCT and similar designs

Despite these limitations it is clear that an experimental method, preferably the full RCT, would have been the evaluation design of choice for many researchers. For example, Denman and co-workers explained that they were unable to complete a fully randomised trial because of the settings in which the research was conducted. In the case of the evaluation of a therapy service for adults who had suffered sexual abuse as children (Smith *et al*) the reasons for not assigning patients to treatment and non-treatment groups were explained in terms of impracticality and uncertain ethics.

Tudor-Smith and co-workers (evaluation of health education programme on substance abuse) stated that they were unable to demonstrate cause and effect because they were unable to use an experimental design. This can be contrasted with the views of van Teijlingen and co-workers (Smokebusters Club evaluation), who, although initially intending to use such a design and following this to a large extent, still concluded that there were far too many variables to allow conclusions to be drawn about cause and effect. The Smokebusters Club activities were, however, far more comprehensive than a single session health education session.

In their non-experimental study of baby feeding practices in a maternity unit Bruce and Griffioen justified the decision not to use a control hospital in terms of resource implications and the inability to influence developments in another setting.

The study by McKenna and co-workers noted other limitations of purely quantitative approaches: they acknowledged that they did not use either open-ended questions in their surveys, or the ethnographic methods, which could have provided richer data.

7.2.2.4 Surveys of patient satisfaction

A number of studies also used surveys to generate views on service reorganisations (Bruce and Griffioen, Gillam *et al,* Smith *et al,* Spiers and Jewell), though doubts have been expressed about these and similar satisfaction surveys that are not grounded in previous qualitative studies of patients' values and aspirations. The pilot study of HIV testing in a community by Chrystie and co-workers is unusual in that it purported to represent patient views, but in this case the questionnaires seem to have been completed by midwives.

7.2.2.5 Generalisability

Despite the apparent belief that results from RCTs can be generalised, this is unfounded: both practitioners and patients may be atypical and patients taking part in a trial may be better treated regardless of the particular group to which they are allocated (Black, 1996). Although having found that their fitness programme had some success, Frost and co-workers proposed that their own (RCT) evaluation for patients with low back pain should be repeated in other centres.

Claims of generalisability are less likely to made of other designs – for example, the **modified action research** reported by McKenna and co-workers who doubted whether the findings from any action research project can be generalised. In considering whole service evaluations, Bruce and Griffioen contended that even if a particular policy has been assessed elsewhere there is no guarantee that application in another locality is going to lead to similar results. Implementation is always shaped by local interests and circumstances, which can of course be explored using qualitative methods.

It may be preferable to concentrate on the potential transferability of a programme, rather than generalisability.[24] Mathews and colleagues were attempting to develop a national resource (in South Africa) in their formative evaluation of an AIDS education programme. In the first school in which this was introduced there were implementation problems. These were explored, and resolved, using further qualitative interviews and the programme was said to be undergoing assessment in another setting.

7.2.3 Strengths of qualitative approaches

The value of qualitative methods in **explanation** has already been noted above in the Jones and Mullee study (i.e. in identifying probable reasons for the lack of significant differences between control and study general practices). Explanations for the relapse from an 'exercise on prescription' scheme emerged from the qualitative part of Lord and Green's study, together with suggestions for improvement. Qualitative methods also have the potential to generate alternative explanations, though within this sample no example has been identified. However, Dujardin and co-workers considered this as a possibility in their study of ante-natal referral compliance: despite testing seven hypotheses they regarded the research as too general to identify the real explanations for low compliance and suggest qualitative research as the next step.

Qualitative methods are also particularly useful in providing **detailed analysis**; for example, the Newens and McEwan study of two sex education programmes. Few of the evaluations examined were purely **process** evaluations, another area where qualitative methods are of value, though the study by Wellard and co-workers carefully outlined the process by which an evaluation system was introduced. This could also be regarded as a **formative** evaluation, as described also by Mathews and co-workers in their development of an AIDS education programme. Hudelson and co-workers reported the use of focused ethnography as a precursor to the development of effective care plans for the treatment of acute respiratory infections. Qualitative methods were also of value in Mathews' research to study programme **implementation**. Implementation issues are equally important in the pilot study reported by Chrystie and co-workers. Given the necessity of introducing HIV testing into community based antenatal care it is clearly essential that this should be accommodated in a way that is acceptable to women (and staff). It is unlikely that this could be achieved without qualitative study, though as noted previously this paper is vague in its descriptions of method.

Qualitative methods are also able to explore and accommodate **different versions of outcome**, taking into account different stakeholder views.

[24] See section 4.1.1.

Many of these studies espoused participatory, or pluralistic, approaches (e.g. Spiers and Jewell, Paterson and Peacock, Lord and Green, van der Walt and Mathews and Wellard).

7.2.4 Failures of qualitative design

The issue of judging quality in qualitative research[25] appears elsewhere in this review and will not, specifically, be dealt with here except to note the sketchiness of some descriptions of method.[26] It seems to be comparatively rare for researchers to acknowledge difficulties with qualitative methods in ways resembling those of quantitative researchers explaining, say, their non-use of the RCT. They may, however, provide a critique of their overall evaluation design. Two papers, with Mathews as co-author in each case (Mathews *et al,* van der Walt and Mathews) examined some of the deficiencies in qualitative techniques (or researchers). In the first case interviews with teachers failed to take account of the strength of religious values, which subsequently impeded the implementation of the AIDS education programme. Further qualitative work did, however, provide a means of problem resolution. The second study has been referred to earlier – significant political actors had been excluded from an evaluation of community health workers. The authors recommended that qualitative researchers should not only concentrate on issues in-depth, but also look at the wider context.

7.2.5 Potential for further qualitative research

A familiar conclusion to almost any research report is that more research is needed. This has been specifically identified as qualitative work in some instances. Jones and Mullee in their rather inconclusive investigation of nurse-led asthma care suggest qualitative interviews with key informants and/or patient focus groups to evaluate patients' perceptions of pro-active care. Having analysed over 5000 ante-natal records and found few explanations for the tendency of women to ignore recommendations about preferred place of delivery, Dujardin and co-workers suggested that qualitative work would be the next step in analysing the problem.

Qualitative methods appear to be the implied method of choice if other research projects were to be continued. Leese and Bosanquet used a five-point scale to record GPs' views about the 1990 contract, the results suggesting some conflict between professional identity and public interest. Any further exploration of this area would require a more qualitative approach. Lindholm and co-workers proposed seeking ways to communicate health messages that took account of individual patient needs, and therefore potentially for qualitative methods. The Smokebusters Club evaluation (van Teijlingen *et al*) omitted to find out what membership of the club actually meant – again, an aspect particularly suitable for qualitative study.

A number of the studies reviewed here make references to the policy applications of evaluations and/or to managerial requirements for information (e.g. Spiers and Jewell, Jones and Mullee, Twinn and Shiu). Policy making on a wider scale does imply that evaluations should be published even when the results are negative or disappointing to programme sponsors and supporters – positive and negative results may be equally important. As Bennett (1995) proposed, there should be an increased publication of innovative evaluations, and the constraints, interpersonal conflicts and management difficulties and failures, which frequently accompany and confuse evaluations, need to be discussed more openly. Thorough and useful evaluation of this type demands qualitative work.

7.2.6 Conclusion

This part of the evaluation case study has demonstrated the diversity and usefulness of qualitative methods currently employed in health service evaluations and made suggestions in relation to further applications. The purpose was to study method, a focus that almost inevitably has pointed to some of the deficiencies in this area, while being less sensitive to potential merit in the complete publications. However, if we are to make the best use of evaluations – if we can learn from them, draw conclusions, and transfer this knowledge to other settings – it becomes incumbent on researchers to state more clearly the reasons why particular methods have been adopted, to be systematic and thorough in their use and to report their work both fully and reflexively.

7.3 Links between the case study and the full report

- This case study demonstrates how the theoretical and philosophical debates outlined in previous sections of the report have impacted

[25] See chapter 5; [26] See section 5.3.3.

upon the development of one area – that of programme evaluation.
- It also demonstrates the historical development of qualitative research in relation to one particular field and shows how the positions adopted by current practitioners often reflect historical controversies.
- It illustrates the debate about the relative merits of qualitative and quantitative research, and the possibility of combining these approaches, to one field.
- It suggests that technologies are always embedded in the context through which they are delivered and that over-reliance upon experimental methods may lead to a neglect of such contexts.
- It illustrates the importance of recognising that, in any area, there are likely to be a number of different perspectives in relation to any innovation or practice and the importance of capturing this range in evaluative research.
- It underlines the importance of relating the choice of method to the current state of knowledge about the phenomenon under investigation.
- It clarifies the range of approaches to evaluation which are currently practised health-related research.
- It illustrates the range of ways in which qualitative and quantitative methods may be combined in the study of a particular programme.
- It illustrates some of the strengths of qualitative approaches, the shortcomings of exclusively quantitative designs and some of the problems that may arise in both qualitative and quantitative evaluations.

Acknowledgements

The authors would like to express their thanks to numerous friends, colleagues and fellow professionals for the interest and encouragement they showed during the writing of this report. In particular, we are grateful to Elizabeth Chisholm and Katherine Clegg for their assistance with section 4.4 and appendix 1, respectively. We are also grateful to Susan Heastor, Social Sciences Librarian at the University of Nottingham, for her invaluable help with bibliographic searching, and to Michelle Taylor who assisted with the consolidation and checking of the bibliography. Constructive feedback from a number of anonymous assessors was also helpful to us.

This review is the product of a collaboration between the five authors. One author took primary responsibility for each section of the report, producing a draft which was then revised in the light of comments from the other authors. Robert Dingwall was responsible for chapter 2, Elizabeth Murphy for chapters 3, 4 and 5, David Greatbatch for chapter 6 and Pamela Watson for chapter 7. Sections 4.4 and 4.5 were drafted by David Greatbatch and Susan Parker, respectively.

References

Aamodt AM (1983). Problems in doing research: developing criteria for evaluating qualitative research. *West J Nurs Res*;**5**:398–401.

Adler P (1985). Wheeling and dealing: an ethnography of an upper-level drug dealing and smuggling community. New York: Columbia University Press.

Adorno TW, Albert H, Dahrendorf R, Habermas J, Pilot H, Popper K (1976). The positivist dispute in German sociology. London: Heinemann.

Agar M (1980). The professional stranger: an informal introduction to ethnography. New York: Academic Press Inc.

Agar M (1991). The right brain strikes back. In: Fielding NG, Lee RM, editors. Using computers in qualitative research. London: Sage. p. 181–94.

Albaek E (1995). Policy evaluation: design and utilization. In: Rist RC, editor. Policy evaluation: linking theory to practice. Aldershot: Elgar. p. 5–18.

Alkin MC, editor (1990). Debates on evaluation. Newbury Park, CA: Sage.

Altheide D, Johnson J (1994). Criteria for assessing interpretive validity in qualitative research. In: Denzin N, Lincoln Y, editors. Handbook of qualitative research. Thousand Oaks, CA: Sage. p. 485–99.

Anderson JG, Aydin CE, Jay SJ, editors (1994a). Evaluating health care information systems: methods and applications. London: Sage.

Anderson JG, Aydin CE, Jay SJ (1994b). Preface. In: Anderson JG, Aydin CE, Jay SJ, editors. Evaluating health care information systems: methods and applications. London: Sage. p. vii–x.

Armstrong D (1995). Finding a 'role' for The Ethnograph in the analysis of qualitative data. In: Burgess RG, editor. Computing and qualitative research. Greenwich CT: JAI Press. p. 63–79.

Aron R (1970). Main currents in sociological thought. Harmondsworth: Penguin.

Athens L (1984). Scientific criteria for evaluating qualitative studies. In: Denzin NK, editor. Studies in symbolic interaction 5. Greenwich, CT: JAI. p. 259–68.

Atkinson JM (1978). Discovering suicide: studies in the social organization of sudden death. London: Macmillan.

Atkinson JM, Heritage J, editors (1984). Structures of social action: studies in conversation analysis. Cambridge: Cambridge University Press.

Atkinson P (1979). Research design in ethnography. In: DE304 Course Team, research methods in education and the social sciences. Milton Keynes: Open University Press. p. 41–82.

Atkinson P (1981). The clinical experience: the construction and reconstruction of medical reality. Aldershot: Gower.

Atkinson P (1983). The reproduction of the professional community. In: Dingwall R, Lewis PSC, editors. The sociology of the professions: lawyers, doctors and others. London: Macmillan. p. 224–41.

Atkinson P, Hammersley M (1994). Ethnography and participant observation. In: Denzin N, Lincoln Y, editors. Handbook of qualitative research. Thousand Oaks, CA: Sage. p. 248–61.

Atkinson P, Coffey A (1997). Analysing documentary realities. In: Silverman D, editor. Qualitative research: theory, methods and practice. London: Sage. p. 45–62.

Atwood J, Hinds P (1986). Heuristic heresy: application of reliability and validity criteria to products of grounded theory. *West J Nurs Res*;**8**:35–47.

Aydin CE (1989). Occupational adaptation to realized medical information systems. *J Health Soc Behav*;**30**: 163–79.

Aydin CE, Ischar R (1989). The effects of computerised order entry on communication between pharmacy and nursing. In: Kingsland LC, editor. 3rd edition. Proceedings of the 13th Annual Symposium on Computer Applications in Medical Care. Washington, DC: IEEE Computer Society. p. 796–801.

Bacon F (1858). The works of Francis Bacon, Baron of Verulam, Viscount St. Alban and Lord High Chancellor of England. London: Longman & Co.

Baker C (1984). The search for adultness: membership work in adolescent–adult talk. *Hum Stud*;**7**:301–23.

Baldamus W (1972). The role of discoveries in social science. In: Shanin T, editor. The rules of the game: cross-disciplinary essays on models in scholarly thought. London: Tavistock. p. 276–302.

Ball S (1981). Beachside Comprehensive. Cambridge: Cambridge University Press.

Barker C (1995). Research and the health services manager in the developing world. *Soc Sci Med*;**41**:1655–65.

Barker P (1995). The ghost road. London: Penguin.

Barnes J (1979). Who should know what? Social science, privacy and ethics. Harmondsworth: Penguin.

Barrett RJ (1988). Clinical writing and the documentary construction of schizophrenia. *Culture, Meaning and Psychiatry*;**12**:265–99.

Bartlett FC (1932). Remembering: a study in experimental and social psychology. Cambridge: Cambridge University Press.

Baruch G (1981). Moral tales: parents' stories of encounters with the health profession. *Sociology of Health and Illness*;**3**:275–96.

Baum F (1995). Researching public health: behind the qualitative quantitative methodological debate. *Soc Sci Med*;**40**:459–68.

Beattie A (1995). Evaluation in community development for health: an opportunity for dialogue. *Health Educ J*; **54**:465–72.

Beauchamp T, Faden R, Wallace RJ, Jnr, Walters L (1982). Introduction. In: Beauchamp T, Faden R, Wallace R, editors. Ethical issues in social science research. Baltimore: Johns Hopkins. p. 3–39.

Beck CT (1993). Qualitative research: the evaluation of its credibility, fittingness and auditability. *West J Nurs Res*;**15**:263–6.

Becker HS (1953). Becoming a marijuana user. *Am J Sociol*;**59**:235–42.

Becker HS (1958). Problems of inference and proof in participant observation. *Am Sociol Rev*;**23**:652–60.

Becker HS, Geer B (1960). Participant observation: the analysis of qualitative data. In: Adams R, Preiss J, editors. Human organization research: field relations and techniques. Homewood, Ill: Dorsey. p. 267–89.

Becker HS, Geer B, Hughes EC, Strauss AL (1961). Boys in white. Chicago: University of Chicago Press.

Becker HS (1967a). Whose side are we on? *Social Problems*;**14**:239–48.

Becker HS (1967b). History, culture and subjective experience: an exploration of the social bases of drug-induced experiences. *J Health Soc Behav*;**8**:163–76.

Becker HS, Geer B (1969a). Participant observation and interviewing: a comparison. In: McCall G, Simmons JC, editors. Issues in participant observation: a text and reader. Reading, MA: Addison-Wesley. p. 322–31.

Becker HS, Geer B (1969b). A rejoinder. In McCall G, Simmons J, editors. Issues in participant observation: a text and reader. Reading, MA: Addison-Wesley. p. 338–41.

Becker HS (1970). Sociological work: method and substance. Chicago: Aldine.

Beiser FC (1987). The fate of reason: German philosophy from Kant to Fichte. Cambridge, MA: Harvard University Press.

Bell C (1977). Reflections on the Banbury Restudy. In: Bell C, Newby H, editors. Doing sociological research. London: George Allen & Unwin. p. 47–62

Bennett F (1995). Qualitative and quantitative methods: in depth or rapid assessment? *Soc Sci Med*;**40**:1589–90.

Berk R (1995). Publishing evaluation research. *Contemp Sociol*;**24**:9–12.

Berkeley G (1967). The principles of human knowledge with other writings. London: Fontana.

Berkeley G (1901). Siris: a chain of philosophical reflexions and inquiries concerning the virtues of tarwater and divers other subjects connected together and owing one from another. In: Fraser EC, editor. The works of George Berkeley, DD, formerly Bishop of Cloyne including his posthumous works. Oxford: Clarendon Press.

Betts N (1996). Recommendations for planning and reporting focus group research. *J Nutr Educ*;**28**(5):279–81.

Bhaskar R (1975). A realist theory of science. Leeds: Leeds Books.

Bhaskar R (1979). The possibility of naturalism: a philosophical critique of the contemporary human sciences. Brighton: Harvester.

Bhaskar R (1993). Dialectic: the pulse of freedom. London: Verso.

Bittner E (1973). Objectivity and realism in sociology. In: Psathas G, editor. Phenomenological sociology. New York: Wiley. p. 109–25.

Black N (1995). The relationship between evaluative research and audit. *J Public Health Med*;**14**:361–6.

Black N (1996). Why we need observational studies to evaluate the effectiveness of health care. *BMJ*;**312**:1215–18.

Blaikie N (1993). Approaches to social enquiry. 1st edition. Cambridge: Polity Press.

Blakey V, Frankland J (1995). Evaluating HIV prevention for women prostitutes in Cardiff. *Health Educ J*;**54**:131–42.

Blomberg J (1995). Ethnography: aligning field studies and system design. In: Monkland A, Gilbert N, editors. Perspectives on human–computer interaction. London: Academic Press. p. 199–222.

Bloor M (1976). Bishop Berkeley and the adeno-tonsillectomy enigma: an exploration of variation in the social construction of medical disposals. *Sociology*;**10**(1):43–61.

Bloor M (1983). Notes on member validation. In: Emerson RM, editor. Contemporary field research: a collection of readings. Boston: Little, Brown. p. 156–72.

Bloor M (1994). On the conceptualization of routine medical decision-making: death certification as a habitual activity. In: Bloor M, Taraborrelli P, editors. Qualitative studies in health and medicine. Aldershot: Avebury. p. 96–109.

Bloor M (1997). Techniques of validation in qualitative research: a critical commentary. In: Miller G, Dingwall R, editors. Context and method in qualitative research. London: Sage. p. 37–50.

Blumer H (1969). Symbolic interactionism. Englewood Cliffs, NJ: Prentice Hall.

Boden D (1994). The business of talk: organizations in action. Cambridge: Polity Press.

Boden D, Zimmerman D, editors (1991). Talk and social structure: studies in ethnomethodology and conversation analysis. Cambridge: Polity Press.

Bogdan R, Biklen S (1982). Qualitative research for education: an introduction to theory and methods. Boston: Allyn & Bacon.

Bosk CL (1992). All God's mistakes: genetic counselling in a pediatric hospital. Chicago: University of Chicago Press.

Bowler I (1997). Problems with interviewing: experiences with service providers and clients. In: Miller G, Dingwall R, editors. Context and method in qualitative research. London: Sage. p. 66–76.

Brewster Smith M (1979). Some perspectives on ethical/political issues in social science research. In: Wax M, Cassell J, editors. Federal regulations: ethical issues and social research. Boulder, Colorado: Westview Press. p. 11–22.

Britten N (1995). Qualitative research: qualitative interviews in medical research. *BMJ*;**311**:251–4.

Brown GW, Harris T (1978). Social origins of depression: a study of psychiatric disorder in women. London: Tavistock.

Bruce N, Griffioen A (1995). Usefulness of a non-experimental study design in the evaluation of service developments for infant feeding in a general hospital. *Soc Sci Med*;**40**(8):1109–16.

Bryman A (1988). Quantity and quality in social research. London: Unwin Hyman.

Bryman A, Burgess RG, editors (1994). Analysing qualitative data. London: Routledge.

Buckholdt D, Gubrium J (1979). Caretakers: treating emotionally disturbed children. Beverley Hills, CA: Sage.

Bulmer M (1979). Concepts in the analysis of qualitative data. In: Bulmer M, editor. Sociological research methods. London: Macmillan.

Bulmer M (1980). Comment on the ethics of covert methods. *Br J Sociol*;**31**:59–65.

Bulmer M (1982). Social research ethics. London: Macmillan.

Bulmer M (1983). The methodology of The Taxi Dance Hall: an early account of Chicago ethnography from the 1920s. *Urban Life*;**12**:95–120.

Bulmer M (1984). The Chicago School of Sociology: institutionalization, diversity and the rise of sociological research. Chicago: University of Chicago.

Bulmer M (1987). Governments and social science science: patterns of mutual influence. In: Bulmer M, editor. Social science research and government. comparative essays in Britain and the United States. Cambridge: Cambridge University Press. p. 1–23.

Burgess R (1982a). Approaches to field research. In: Burgess R, editor. Field research: a sourcebook and field manual. London: George Allen & Unwin. p. 1–11.

Burgess RG (1982b). Some role problems in field research. In: Burgess RG, editor. Field research: a sourcebook and field manual. London: George Allen & Unwin. p. 45–9.

Burgess RG (1984). In the field: an introduction to field research. London: Unwin Hyman.

Burgess R (1985). The whole truth? Some ethical problems of research in a comprehensive school. In: Burgess RG, editor. Strategies of educational research: qualitative methods. Lewes: Falmer. p. 141–62.

Burke P (1981). Montaigne. Oxford: Oxford University Press.

Burns N, Grove S (1993). The practice of nursing research: conduct, critique and utilization. Philadelphia: WB Saunders.

Burton L (1975). The family life of sick children. London: Routledge.

Button G, Lee J, editors (1987). Talk and social organisation. Avon, England: Multilingual Matters.

Campbell D (1957). Factors relevant to the validity of experiments in social settings. *Psychol Bull*;**54**:297–312.

Campbell D, Fiske D (1959). Convergent and discriminant validation by multitrait multi-dimensional matrix. *Psychol Bull*;**56**:81–105.

Campbell D (1978). Qualitative knowing in action research. In: Brenner M, Marsh P, Brenner M, editors. The social contexts of method. London: Croom Helm. p. 184–209.

Campbell D (1979). "Degrees of Freedom" and the case study. In: Cook TD, Reichardt CS, editors. Qualitative and quantitative methods in evaluation research. Beverley Hills, CA: Sage. p. 49–67.

Campbell D (1984). Can we be scientific in applied social science? In: Conner R, Altman D, Jackson C, editors. Evaluation studies review annual volume 9. Beverley Hills, CA: Sage. p. 26–48.

Campbell D (1994). Can we overcome world-view incommensurability/relativity in trying to understand the other? In: Jessor R, Colby A, Shweder R, editors. Ethnography and human development: context and meaning in social inquiry. Chicago: University of Chicago Press. p. 153–72.

Caradog Jones D (1949). Social surveys. London: Hutchinson.

Carey JT (1975). Sociology and public affairs: The Chicago School. Beverley Hills, CA: Sage.

Cash JI, Lawrence PR, editors (1989). The information systems research challenge: qualitative research methods. Harvard Business School Research Colloquium, vol. 1. Boston, MA: Harvard Business School.

Cassell J (1978). Risk and benefit to subjects of fieldwork. *Am Sociol*;**13**:134–43.

Cassell J (1979). Regulating fieldwork: of subjects, subjection and intersubjectivity. In: Wax M, Cassell J, editors. Federal regulations: ethical issues and social research. Boulder, Colorado: Westview Press. p. 129–44.

Cassell J, Wax M (1980). Editorial introduction: toward a moral science of human beings. *Social Problems*;**27**:259–64.

Cassell J (1982). Does risk–benefit analysis apply to moral evaluation of social research? In: Beauchamp T, Faden R, Wallace R, Walters L, editors. Ethical issues in social science research. Baltimore: Johns Hopkins University Press.

Chalmers A (1982). What is this thing called science? St Lucia, QLD: University of Queensland Press.

Charmaz K (1983a). The grounded theory method: an explication and interpretation. In: Emerson RM, editor. Contemporary field research: a collection of readings. Prospect Heights, Ill: Waveland Press. p. 109–26.

Charmaz K (1983b). Loss of self: a fundamental form of suffering in the chronically ill. *Sociology of Health and Illness*;**5**:168–95.

Chrystie I, Wolfe C, Kennedy J, Zander I, Tilzey A, Banatvala J (1995). Voluntary, named testing for HIV in a community-based antenatal clinic. *BMJ*;**311**:928–31.

Cicourel AV (1964). Method and measurement in sociology. New York: Free Press.

Cicourel AV (1968). The social organization of juvenile justice. New York: John Wiley & Son.

Cicourel AV (1974). Theory and method in a study of Argentine fertility. New York: Wiley.

Cicourel AV, Kitsuse JI (1963). The educational decision-makers. Indianapolis: Bobbs-Merill.

Clarke L (1995). Nursing research: science, visions and telling stories. *JAdv Nurs*;**21**:584–93.

Clavarino A, Najman J, Silverman D (1995). Assessing the quality of qualitative data. *Qualitative Inquiry*;**1**:223–42.

Clayman SE, Maynard D (1995). Ethnomethodology and conversation analysis. In: ten Have P, Psathas G, editors. Situated order: studies in the social organisation of talk and embodied activities. Lanham, MD: University Press of America. p. 1–30.

Cochrane AL (1972). Effectiveness and efficiency: random reflections on health services. London: Nuffield Provincial Hospitals Trust.

Cockburn D, Bourne G (1983). Reason and experience: Hume. Milton Keynes: Open University Press.

Coffey A, Holbrook B, Atkinson P (1996). Qualitative data analysis: technologies and representations. *Sociological Research Online*;**1**(1).

Cohen S, editor (1971). Images of deviance. Harmondsworth: Penguin.

Collingwood RG (1946). The idea of history. Oxford: Clarendon Press.

Comte A (1970). Introduction to positive philosophy. Indianapolis: Bobbs-Merrill.

Conrad P, Reinharz S (1984). Computers and qualitative data: editor's intoductory essay. *Qualitative Sociology*;**7**:1–2.

Cook T, Campbell D (1979). Quasi-experimentation. Chicago: Rand McNally.

Cook TD, Reichardt CS, editors (1979). Qualitative and quantitative methods in evaluation research. Beverley Hills, CA: Sage.

Cornwell J (1984). Hard-earned lives: accounts of health and illness from East London. London: Tavistock.

Coulter A (1991). Evaluating the outcomes of health care. In: Gabe J, Calnan M, Bury M, editors. The sociology of the health service. London: Routledge. p. 115–39.

Cowman S (1993). Triangulation: a means of reconciliation in nursing research. *JAdv Nurs*;**18**:788–92.

Coxon APM (1988). Something sensational....the sexual diary as a tool for mapping detailed sexual behaviour. *Sociol Rev*;**36**(2):253–67.

Crabtree BF, Miller WL (1991). A qualitative approach to primary care research: the long interview. *Family Medicine*;**23**:145–51.

Cressey D (1950). The criminal violation of financial trust. *Am Sociol Rev*;**15**:738–43.

Cressey PG (1983). A comparison in the roles of the 'sociological stranger' and the 'anonymous stranger' in field research. *Urban Life*;**12**(1):102–20.

Cronbach LJ (1975). Beyond the two disciplines of scientific psychology. *Am Psychol*;**30**:127.

Cronbach LJ, Associates (1980). Toward reform of program evaluation. San Francisco: Jossey Bass.

Crow G, Allan G (1994). Community life: an introduction to local social relations. London: Harvester Wheatsheaf.

Cunningham-Burley S (1984). We don't talk about it: issues of gender and method in the portrayal of grand-fatherhood. *Sociology*;**18**:325–38.

Davis F (1963). Passage through crisis: polio victims and their families. Indianapolis: Bobbs-Merill.

Davies RM, Atkinson PA (1991). Students of midwifery: 'doing the obs' and other copying strategies. *Midwifery*;**7**:113–21.

Dean J, Eichhorn R, Dean L (1969). Limitations and advantages of unstructured methods. In: McCall G, Simmons J, editors. Issues in participant observation: a text and reader. Reading, MA: Addison-Wesley. p. 19–24.

Delamont S, Hamilton D (1976). Classroom research: a critique and new approach. In: Stubbs M, Delamont S, editors. Explorations in classroom observation. London: John Wiley & Sons. p. 3–20

Denman S, Pearson J, Moody D, Davies P, Madeley R (1995). Theatre in education on HIV and AIDS: a controlled study of schoolchildren's knowledge and attitudes. *Health Educ J*;**54**:3–17.

Denscombe M (1983). Interviews, accounts and ethnographic research on teachers. In: Hammersley M, editor. The ethnography of schooling. Driffield: Nafferton Books. p. 107–28.

Denzin N (1970). The research act. 1st edition. Englewood Cliffs, NJ: Prentice Hall.

Denzin N (1971). The logic of naturalistic inquiry. *Social Forces*;**50**:166–82.

Denzin N (1978). The research act. 2nd edition. New York: McGraw-Hill.

Denzin N (1983). Interpretive interactionism. In: Morgan D, editor. Beyond method: strategies for social research. Beverley Hills, CA: Sage. p. 129–46.

Denzin N, Lincoln YS (1994a). Introduction to Part V, The art of interpretation, evaluation and presentation. In: Denzin N, Lincoln Y, editors. Handbook of qualitative research. Thousand Oaks, CA: Sage. p. 479–83.

Denzin N, Lincoln Y (1994b). Entering the field of qualitative research. In: Denzin NK, Lincoln YS, editors. Handbook of qualitative research. Thousand Oaks, CA: Sage. p. 1–17.

DeVault M (1990). Talking and listening from women's standpoint: feminist strategies for interviewing and analysis. *Social Problems*;**37**:96–116.

de Vries H, Weijts W, Dijkstra M, Kok G (1992). The utilization of qualitative and quantitative data for health education program planning, implementation and evaluation: a spiral approach. *Health Educ Q*;**19**:101–15.

Dey I (1993). Qualitative data analysis: a user-friendly guide for social scientists. London: Routledge.

Dibble VK (1975). The legacy of Albion Small. Chicago: University of Chicago Press.

Diener E, Crandall R (1978). Ethics in social and behavioural research. Chicago: University of Chicago Press.

Dilman I (1973). Induction and deduction: a study in Wittgenstein. Oxford: Basil Blackwell.

Dilorio C (1994). Focus groups: an interview method for nursing research. *J Neurosci Nurs*;**26**(3):175–80.

Dingwall R (1980). Ethics and ethnography. *Sociol Rev*;**28**:871–91.

Dingwall R (1981). The ethnomethodological movement. In: Payne G, Dingwall R, Payne J, Carter M, editors. Sociology and social research. London: Croom Helm. p. 124–38.

Dingwall R, Murray T (1983). Categorization in accident departments; 'good' patients, 'bad' patients and 'children'. *Sociology of Health and Illness*;**5**:127–48.

Dingwall R (1992). Don't mind him – he's from Barcelona: qualitative methods in health studies. In: Daly J, McDonald I, Wilks E, editors. Researching health care. London: Tavistock/Routledge. p. 161–75.

Dingwall R (1997a). Accounts, interviews and observations. In: Miller G, Dingwall R, editors. Context and method in qualitative research. London: Sage. p. 51–65.

Dingwall R (1997b). The moral discourse of interactionism. In: Miller G, Dingwall R, editors. Context and method in qualitative research. London: Sage. p. 198–205.

Dollard J (1949). Caste and class in a southern town. New York: Harper & Brothers.

Dootson S (1995). An in-depth study of triangulation. *J Adv Nurs*;**22**:183–7.

Douglas J (1967). The social meanings of suicide. Princeton, NJ; Princeton University Press.

Douglas J (1976). Investigative social research: individual and team field research. Beverley Hills, CA: Sage.

Dowling AF, Jr (1980). Do hospital staff interfere with computer system implementation? *Health Care Manage Rev*;**5**:23–32. Reprinted in Anderson JG, Jay SJ, editors (1987). Use and impact of computers in clinical medicine. New York: Springer Verlag. p. 302–17.

Drew P, Heritage J, editors (1992a). Talk at work: interaction in institutional settings. Cambridge: Cambridge University Press.

Drew P, Heritage J (1992b). Analyzing talk at work: an introduction. In: Drew P, Heritage J, editors. Talk at work. Cambridge: Cambridge University Press. p. 3–65.

Duffy M (1985). Designing nursing research: the qualitative-quantitative debate. *J Adv Nurs*;**10**:225–32.

Duffy M (1987). Methodological triangulation: a vehicle for merging quantitative and qualitative research methods. *Image: Journal of Nursing Scholarship*;**19**:130–3.

Dujardin B, Clarysse G, Criel B, de Brouwere V, Wangata N (1995). The strategy of risk approach in antenatal care: evaluation of the referral compliance. *Soc Sci Med*;**40**(4):529–52.

Dunn J, Urmson JO, Ayer AJ (1992). The British empiricists: Locke, Berkeley, Hume. Oxford: Oxford University Press.

Durkin T (1997). Using computers in qualitative research. In: Miller G, Dingwall R, editors. Context and method in qualitative research. London: Sage. p. 92–105.

Eakin JM, Maclean HM (1992). A critical perspective on research and knowledge development in health promotion. *Can J Public Health*;**83**:572–6.

Easterday L, Papademas D, Schorr L, Valentine C (1982). The making of a field researcher: role problems in fieldwork. In: Burgess R, editor. Field research: a sourcebook and field manual. London: George Allen & Unwin. p. 62–78.

Eisenhart MA, Howe KR (1992). Validity in educational research. In: LeCompte MD, Milroy WL, Preissle J, editors. The handbook of qualitative research in education. New York: Academic Press. p. 643–80.

Eisner E (1983). Anastasia might still be alive, but the monarchy is dead. *Educational Researcher*;**12**.

Emerson R (1981). Observational fieldwork. *Annu Rev Sociol*;**7**:351–78.

Emerson R (1983). Contemporary field research: a collection of readings. Prospect Heights, Ill: Waveland Press.

Emerson RM, Pollner M (1988). On members' responses to researchers' accounts. *Human Organization*;**47**:189–98.

Engel G (1977). The need for a new medical model: a challenge for biomedicine. *Science*;**196**:129–6.

Erikson K (1967). A comment on disguised observation in sociology. *Social Problems*;**14**:366–73.

Eriksson CG (1988). Focus groups and other methods for increased effectiveness of community intervention – a review. *Scand J Prim Health Care Suppl*;**1**:73–80.

Evans J (1983). Criteria of validity in social research: exploring the relationship between ethnographic and quantitative approaches. In: Hammersley M, editor. The ethnography of schooling: methodological issues. Driffield: Nafferton.

Evans-Pritchard EE (1940). The Nuer: a description of the modes of livelihood and political institutions of a Nilotic people. Oxford: Clarendon Press.

Fafchamps D (1991). Ethnographic work flow analysis: specifications for design. In: Bullinger JH, editor. Proceedings of the 4th International Conference on Human Computer Interaction. Amsterdam: Elsevier Science Publishers.

Fafchamps D, Young CY, Tang PC, editors (1991). Modelling work practice: design and evaluation of a physician workstation. Pala Alto, CA: Hewlett-Packard Laboratories.

Faris REL (1967). Chicago sociology (1920–1932). Chicago: University of Chicago Press.

Fehling D (1989). Herodotus and his 'sources': citation, invention and narrative art. Leeds: Francis Cairns.

Festinger L, Riecken H, Schachter S (1956). When prophecy fails. New York: Harper & Row.

Fetterman DM (1989). Ethnography: step by step. Newbury Park, CA: Sage.

Feyerabend P (1975). Against method. London: Verso.

Fielding N (1982). Observational research on the National Front. In: Bulmer M, editor. Social research ethics. London: Macmillan. p. 80–104.

Fielding N, Fielding J (1986). Linking data. London: Collier MacMillan.

Fielding NG, Lee RM, editors (1991). Using computers in qualitative research. London: Sage.

Fielding N (1993). Ethnography. In: Gilbert N, editor. Researching social life. London: Sage. p. 154–71.

Filstead W (1981). Using qualitative methods in evaluation research. An illustrative bibliography. *Evaluation Review*;**5**:259–68.

Finch J (1984). 'It's great to have someone to talk to': the ethics and politics of interviewing women. In: Bell C, Roberts H, editors. Social researching: policies, problems and practice. London: Routledge & Kegan Paul. p. 70–85.

Finch J (1986). Research and policy: the uses of qualitative methods in social and educational research. Lewes: Falmer Press.

Fine GA, editor (1995). A Second Chicago School? The development of a postwar American sociology. Chicago: University of Chicago Press.

Finnegan R, Thomas R (1993). Documentary sources, statistical records and data bases. In: DEH313 Course Team. Principles of social and educational research, Block 3, data collection and construction. Milton Keynes: Open University Press.

Finnegan R (1996). Using documents. In: Sapsford R, Jupp V, editors. Data collection and analysis. London: Sage. p. 138–51.

Firestone W (1993). Alternative arguments for generalizing from data as applied to qualitative research. *Educational Researcher*;**22**:16–23.

Fischer PJ, Stratmann WC, Steele DJ (1980). User reaction to PROMIS: issues related to acceptability of medical innovations. In: O'Neill JT, editor. Proceedings of the 4th Annual Symposium on Computer Applications in Medical Care. Silver Spring, MD: IEEE Computer Society Press. p. 1722–30. Reprinted in: Anderson JG, Jay SJ, editors (1987). Use and impact of computers in clinical medicine. New York: Springer Verlag. p. 284–301.

Fletcher R (1971a). The making of sociology: a study of sociological theory: Volume 1 Beginnings and foundations. London: Michael Joseph.

Fletcher R, editor (1971b). John Stuart Mill: a logical critique of sociology. London: Michael Joseph.

Flick U (1992). Triangulation revisited: strategy of validation or alternative. *J Theory Soc Behav*;**22**(2):175–97.

Fontana A, Frey J (1994). Interviewing: the art of science. In: Denzin N, Lincoln Y, editors. Handbook of qualitative research. Thousand Oaks, CA: Sage. p. 361–76.

Forsythe S, Buchanan BG (1991). Broadening our approach to evaluating medical expert systems. In: Clayton PD, editor. Proceedings of the 15th Annual Symposium on Computer Applications in Medical Care. New York: McGraw-Hill. p. 8–12.

Foster P (1993). Unit 12 observational research. In: DEH313 Course Team. Principles of social and educational research. Milton Keynes: Open University Press. p. 37–73.

Fox R (1974). Experiment perilous: physicians and patients facing the unknown. Pittsburgh, PA: University of Pennsylvanial Press.

Frake C (1962). The ethnographic study of cognitive systems. In: Gladwin T, Sturtevant W, editors. Anthropology and human behaviour. Washington DC: Anthropological Society of Washington.

Frake C (1964). How to ask for a drink in Subanun. *Am Anthropol*;**66**:127–32.

Frake C (1975). How to enter a Yakan house. In: Sanchez M, Blount B, editors. Socio-culture dimensions of language use. New York: Academic. p. 25–40.

Frankenberg R (1957). Village on the border. London: Cohen and West.

Frankenberg R (1966). Communities in Britain: social life in town and country. Harmondsworth: Penguin.

Frankenberg R, editor (1992). Time, health and medicine. London: Sage.

Frazer JG (1890). The golden bough. London: Macmillan.

Freeman H, Solomon M (1981). The next decade in evaluation research. In: Levine RA, Solomon MA, Hellstern G, Wollman H, editors. Evaluation research and practice. Comparative and international perspectives. Beverley Hills, CA: Sage. p. 12–26.

Freund J (1968). The sociology of Max Weber. London: Allen Lane.

Friedman N (1967). The social nature of psychological research. New York: Basic Books.

Friedson E (1970). Profession of medicine: a study in the sociology of applied knowledge. New York: Dodd Mead.

Frost H, Klaber Moffett J, Moser J, Fairbank J (1995). Randomized controlled trial for evaluation of fitness programme for patients with low back pain. *BMJ*;**310**:151–4.

Galliher J (1973). The protection of human subjects: a reexamination of the professional code of ethics. *Am Sociol*;**8**:93–100.

Galliher J (1980). Social scientists' ethical responsibilities to superordinates: looking upward meekly. *Social Problems*;**27**:298–308.

Gans H (1982). The participant observer as a human being: observations on the personal aspects of fieldwork. In: Burgess RG, editor. Field research: a sourcebook and field manual. London: George Allen & Unwin. p. 53–61.

Gantley M, Davies D, Murcott A (1993). Sudden infant death syndrome: links with infant care practices. *BMJ*;**306**:16–20.

Garfinkel H (1967). Studies in ethnomethodology. Englewood Cliffs, NJ: Prentice-Hall.

Gee W, editor (1929). Research in the social sciences. London: Macmillan.

Geertz C (1973). The interpretation of culture. New York: Basic Books.

Gerson E (1984). Qualitative research and the computer. *Qualitative Sociology*;**7**:61–74.

Gibson H (1978). The dressing clinic in an accident and emergency department. In: Davis A, editor. Relationships between doctors and patients. Farnborough: Teakfield.

Giddens A (1974). Introduction. In: Giddens A, editor. Positivism and sociology. London: Heinemann Educational.

Gillam S, Ball M, Prasad M, Dunne H, Cohen S, Vafidis G (1995). Investigation of the benefits and costs of an ophthalmic outreach clinic in general practice. *Br J Gen Pract*;**45**:649–52.

Glaser BG, Strauss A (1965a). The discovery of substantive theory: a basic strategy underlying qualitative research. *Am Behav Scientist*;**8**:5–12.

Glaser BG, Strauss A (1965b). Awareness of dying. Chicago: Aldine Publishing Company.

Glaser BG, Strauss AL (1967). The discovery of grounded theory: strategies for qualitative research. New York: Aldine.

Glaser BG (1978). Theoretical sensitivity. Mill Valley, CA: Sociology Press.

Glaser BG (1992). Emergence vs. forcing: basics of grounded theory analysis. Mill Valley, CA: The Sociology Press.

Glennerster H, Korman N, Marsden-Wilson F (1983). Planning for priority groups. Oxford: Martin Robertson.

Goffman E (1951). Symbols of class status. *Br J Sociol*;**2**:294–304.

Goffman E (1959). The presentation of self in everyday life. Harmondsworth: Penguin.

Goffman E (1961). Asylums. New York: Doubleday.

Goffman E (1983). Felicity's condition. *Am J Sociol*;**83**:1–53.

Gold R (1969). Roles in sociological field observations. In: McCall G, Simmons J, editors. Issues in participant observation: a text and reader. Reading, MA: Addison-Wesley. p. 30–9.

Good MD (1992). Qualitative designs for assessing interventions in primary care: a discussion. In: Tudiver F, Bass MJ, Dunn EV, Norton PG, Stewart M, editors. Assessing interventions. Traditional and innovative methods. Newbury Park, CA: Sage. p. 96–105.

Goodwin C (1981). Conversational organization: interaction between speakers and hearers. New York: Academic Press.

Goodwin C, Goodwin MH (1992). Professional vision. In: Proceedings of International Conference on Discourse and Professions, September.

Goodwin C, Heritage J (1990). Conversation analysis. *Annu Rev Anthropol*;**19**:283–307.

Goodwin C, Goodwin MH (1997). Formulating planes: seeing as a situated activity. In: Middleton Y, Engestrom D, editors. Distributed cognition. London: Sage.

Goodwin L, Goodwin H (1984a). Qualitative vs quantitative or qualitative and quantitative research? *Nurs Res*;**33**:378–80.

Goodwin L, Goodwin W (1984b). Are reliability and validity 'relevant' in qualitative evaluation research? *Evaluation and the Health Professions*;**7**:413–26.

Gordon I, Whitear B, Guthrie D (1997). Stopping them starting: evaluation of a community-based project to discourage teenage smoking in Cardiff. *Health Educ J*;**56**:42–50.

Gould L, Walker A, Crane L, Lidz C (1974). Connections: notes from the heroin world. New Haven: Yale University Press.

Graffam Walker A (1986). The verbatim record: the myth and the reality. In: Fisher S, Todd A, editors. Discourse and institutional authority: medicine, education and law. Norwood, NJ, Ablex. p. 205–22.

Greatbatch D, Luff P, Heath CC, Campion P (1993). Interpersonal communication and human–computer interaction: an examination of the use of computers in medical consultations. *Interacting With Computers*;**5**:193–216.

Greatbatch D, Heath CC, Campion P, Luff P (1995a). How do desk-top computers affect the doctor–patient interaction? *Family Practice*;**12**(1):32–6.

Greatbatch D, Heath CC, Luff P, Campion P (1995b). Conversation analysis: human–computer interaction and the general practice consultation. In: Monk A, Gilbert N, editors. Perspectives on human–computer interaction. London: Academic Press. p. 199–222.

Greene J (1994). Qualitative program evaluation: practice and promise. In: Denzin N, Lincoln Y, editors. Handbook of qualitative research. Thousand Oaks, CA: Sage. p. 530–44.

Greene JC (1996). Qualitative evaluation and scientific citizenship: reflections and refractions. *Evaluation*;**2**:277–89.

Green J, Franquiz M, Dixon C (1997). The myth of the objective transcript: transcribing as a situated act. *TESOL Q*;**31**:172–6.

Greenhouse CJ, Yngvesson B, Engel DM (1994). Law and community in three American towns. Ithaca: Cornell University Press.

Griffen C (1986). Qualitative methods and female experience: young women from school to the job market. In: Wilkinson S, editor. Feminist social psychology: developing theory and practice. Milton Keynes: Open University Press. p. 173–91.

Griffith AI, Smith DE (1987). Constructing cultural knowledge: mothering as discourse. In: Gaskell J, McLaren A, editors. Women and education: a Canadian perspective. Calgary: Detselig. p. 87–103.

Guba EG (1981). Criteria for assessing the trustworthiness of naturalistic inquiries. *Educ Commun Technol J*;**29**:75–91.

Guba EG, Lincoln YS (1989). Fourth generation evaluation. Newbury Park, CA: Sage.

Guba EG (1990). The paradigm dialog. Newbury Park, CA: Sage.

Guba EG, Lincoln YS (1994). Competing paradigms in qualitative research. In: Denzin NK, Lincoln YS, editors. Handbook of qualitative research. Thousand Oaks, CA: Sage.

Gubrium J (1975). Living and dying at Murray Manor. New York: St Martin's.

Gubrium J (1978). Notes on the social organization of senility. *Urban Life*;**7**:23–44.

Gubrium J, Buckholdt D (1982). Describing care: image and practice in rehabilitation. Cambridge MA: Oelgeschlager, Gunn & Hain.

Gubrium J, Buckholdt D, Lynott R (1982). Considerations on a theory of descriptive activity. *Mid American Review of Sociology*;**6**(1):17–35.

Guillemin JH, Holmstrom LL (1986). Mixed blessings: intensive care for newborns. New York: Oxford University Press.

Gussow Z (1964). The observer–observed relationship as information about structure in small-group research: a comparative study of urban elementary school classrooms. *Psychiatry*;**27**:230–47.

Habermann-Little B (1991). Qualitative research methodologies: an overview. *J Neurosci Nurs*;**23**:188–90.

Halfpenny P (1979). The analysis of qualitative data. *Sociol Rev*;**27**:799–825.

Hamilton D, Jenkins D, King C, MacDonald B, Parlett M, editors (1977). Beyond the numbers game: a reader in educational evaluation. Basingstoke: Macmillan Education.

Hammersley M, for the DE304 Course Team (1979a). The process of analysis: generating and testing theory. In: Making sense of data. Milton Keynes: Open University Press. p. 22–30.

Hammersley M, for the DE304 Course Team (1979b). Data collection in ethnographic research. In: Research methods in education and the social sciences: data collection procedures. Milton Keynes: Open University Press. p. 89–175.

Hammersley M (1985). From ethnography to theory: a programme and paradigm in the sociology of education. *Sociology*;**19**:244–59.

Hammersley M, Scarth J, Webb S (1985). Developing and testing theory: the case of research on pupil learning and examinations. In: Burgess RG, editor. Issues in educational research: qualitative methods. London: Falmer Press. p. 48–66.

Hammersley M (1987). Some notes on the terms 'validity' and 'reliability'. *Br Educ Res J*;**13**:73–81.

Hammersley M (1989). The dilemma of qualitative method: Herbert Blumer and the Chicago tradition. London: Routledge.

Hammersley M (1990). Reading ethnographic research. New York: Longman.

Hammersley M (1992a). What's wrong with ethnography? London: Routledge.

Hammersley M (1992b). What's wrong with ethnography? The myth of theoretical description. In: Hammersley M, editor. What's wrong with ethnography? London: Routledge. p. 11–31.

Hammersley M (1992c). Some questions about theory in ethnography and history. In: Hammersley M, editor. What's wrong with ethnography? London: Routledge. p. 32–42.

Hammersley M (1992d). Ethnography and realism. In: Hammersley M, editor. What's wrong with ethnography? London: Routledge. p. 43–56.

Hammersley M (1992e). By what criteria should ethnographic research be judged? In: Hammersley M, editor. What's wrong with ethnography? London: Routledge. p. 57–82.

Hammersley M (1992f). The generalisability of ethnography. In: Hammersley M, editor. What's wrong with ethnography? London: Routledge. p. 85–95.

Hammersley M (1992g). Deconstructing the qualitative–quantitative divide. In: Hammersley M, editor. What's wrong with ethnography? London: Routledge. p. 159–82.

Hammersley M (1993). What is social research? Milton Keynes: Open University Press.

Hammersley M, Atkinson P (1995). Ethnography: principles in practice. London: Routledge.

Harding S (1987). Feminism and methodology. Bloomington: Indiana University Press.

Hargreaves D (1967). Social relations in a secondary school. London: Routledge & Kegan Paul.

Harre R (1970). The principles of scientific thinking. London: Macmillan.

Harre R, Secord PF (1972). The explanation of social behaviour. Oxford: Basil Blackwell.

Harrisson T (1947). The future of sociology. *Pilot Papers*;**2**(1):10–21.

Hawthorn G (1976). Enlightenment and despair: a history of sociology. Cambridge: Cambridge University Press.

Heap JL (1995). Constructionism in the rhetoric and practice of fourth generation evaluation. *Evaluation and Program Planning*;**18**:51–61.

Heath CC (1986). Body movement and speech in medical interaction. Cambridge: Cambridge University Press.

Heath CC (1997). The analysis of activities in face to face interaction using video. In: Silverman D, editor. Qualitative research: theory, method and practice. London: Sage. p. 183–200.

Heath CC, Luff P (1992a). Collaboration and control: crisis management and multimedia technology in London Underground line control rooms. *CSCW J*;**1**:69–94.

Heath CC, Luff P (1992b). Media space and communicative assymetries: preliminary observations of video-mediated interaction. *Human-Computer Interaction*;**7**:315–46.

Heath CC, Luff P (1993). Explicating face-to-face interaction. In: Gilbert N, editor. Researching social life. London: Sage. p. 306–26.

Heath CC, Jirotka M, Luff P, Hindmarsh J (1995). Unpacking collaboration: the interactional organization of trading in a city dealing room. In: Proceedings of The European Computer Supported Co-operative Work, September 13–17. p. 155–70.

Heath CC, Luff P (1996). Documents and professional practice: 'bad' organizational reasons for 'good' clinical records. Working paper. The Work Interaction and Technology Research Group. King's College, University of London.

Heldo H, Wildavsky A (1974). The private government of public money, London: Macmillan.

Henwood K, Pidgeon N (1992). Qualitative research and psychological theorising. *Br J Psychol*;**83**:97–111.

Henwood K, Pidgeon N (1993). Qualitative research and psychological theorizing. In: Hammersley M, editor. Social research: philosophy, politics and practice. London: Sage. p. 14–32.

Heritage JC (1984). Garfinkel and ethnomethodology, Cambridge: Polity Press.

Heritage JC (1988). Explanations as accounts: a conversation analytic perspective. In: Antaki C, editor. Understanding everyday explanation: a casebook of methods. Beverley Hills, CA: Sage. p. 127–44.

Heritage J, Sefi S (1992). Dilemmas of advice: aspects of the delivery and reception of advice in interactions between health visitors and first time mothers. In: Drew P, Heritage J, editors. Talk at work. Cambridge: Cambridge University Press.

Heritage JC (1995). Conversational analysis: methodological aspects. In: Quasthoff UM, editor. Aspects of oral communication. Berlin, Germany and Chicago, Ill: De Gruyter. p. 391–418.

Heritage JC (1997). Conversation analysis and institutional talk. In Silverman D, editor. Qualitative research: theory, method and practice. London: Sage. p. 161–82.

Herndon S, Kreps G (1995). Qualitative research. Applications in organizational communication. Cresskill, NJ: Hampton Press.

Hesse-Biber S (1995). Unleashing Frankenstin's monster? The social system of analytic fieldwork. In: Burgess RG, editor. Computing and qualitative research. Greenwich CT: JAI Press. p. 25–41.

Hilbert R (1980). Covert participant observation: on its nature and practice. *Urban Life*;**9**:51–78.

Hinds P, Scandrett-Hibden S, McAulay L (1990). Further assessment of a method to estimate reliability and validity of qualitative research findings. *J Adv Nurs*;**15**:430–5.

Hirschman AO (1977). The passions and the interests: political arguments for capitalism before its triumph. Princeton, NJ: Princeton University Press.

Holdaway S (1982). 'An inside job': a case study of covert research on the police. In: Bulmer M, editor. Social research ethics: an examination of the merits of covert participant observation. London: Macmillan.

Holland WW, editor (1983). Evaluation of health care. Oxford: Oxford University Press.

Holman HR (1993). Qualitative inquiry in medical research. *J Clin Epidemiol*;**46**:29–36.

Holsti OR (1969). Content analysis for the social sciences and humanities. Reading, MA: Addison Wesley.

Homan R (1980). The ethics of covert methods. *Br J Sociol*;**31**:46–59.

Honigman J (1982). Sampling in ethnographic field-work. In: Burgess RG, editor. Field research: a source-book and field Manual. London: George, Allen & Unwin. p. 79–90.

Hopper R (1991). Ethnography and conversation analysis after talking culture. *Research on Language and Social Interaction*;**24**:161–70.

Hornblower S (1987). Thucydides. London: Duckworth.

House E (1990). An ethics of qualitative field studies. In: Guba E, editor. The paradigm dialog. Newbury Park, CA: Sage. p. 158–201.

Howe K, Eisenhart M (1990). Standards for qualitative (and quantitative) research: a prolegomenon. *Educational Researcher*;**19**:2–9.

Hudelson P, Huanca T, Charaly D, Cirpa V (1995). Ethnographic studies of ARI in Bolivia and their use by the national ARI programme. *Soc Sci Med*;**41**(12):1677–83.

Hughes D (1977). Everyday and medical knowledge in categorising patients. In: Dingwall R, Heath C, Reid M, Stacey M, editors. Health care and health knowledge. London: Croom Helm.

Hughes D (1980). The ambulance journey as an information-generating process. *Sociology of Health and Illness*;**2**:115–32.

Hughes EC (1971). The sociological eye. Chicago: Aldine-Atherton.

Hughes HS (1959). Consciousness and society: the reorientation of European social thought 1890–1930. London: MacGibbon & Kee.

Hughes JA, King V, Rodden T, Andesen H (1994). Moving out of the control room: ethnography in system design. Working paper. Department of Sociology, University of Leicester.

Hume D (1975). Enquiries concerning human understanding and concerning the principles of morals. 3rd edition. Oxford: Oxford University Press.

Hume D (1978). A treatise of human nature. 2nd edition. Oxford: Clarendon Press.

Humphreys L (1970). Tearoom trade. Chicago: Aldine.

Hunt J (1984). The development of rapport through the negotiation of gender in fieldwork among the police. *Human Organization*;**43**:283–96.

Hunter V (1982). Past and process in Herodotus and Thucydides. Princeton: Princeton University Press.

Ianni F, Ianni E (1972). A family business: kinship and social control in organized crime. London: Routledge.

Illsley R (1980). Professional or public health? Sociology in health and medicine. London: Nuffield Provincial Hospitals Trust.

Imle M, Atwood J (1988). Retaining qualitative validity while gaining quantitative reliability and validity: development of the transition to parenthood concerns scale. *Ans-Advances in Nursing Science*;**11**:61–75.

Janesick V (1994). The dance of qualitative research design: methaphor, methodolatory and meaning. In: Denzin N, Lincoln Y, editors. Handbook of qualitative research. Thousand Oaks, CA: Sage. p. 209–19.

Jeffery R (1979). Normal rubbish: deviant patients in casualty departments. *Sociology of Health and Illness*;**1**:90–107.

Jefferys M (1965). An anatomy of social welfare staff and their clients in the county of Buckinghamshire. London: Michael Joseph.

Jennett B (1992). Health technology assessment (editorial). *BMJ*;**305**:67–8.

Jensen G (1989). Qualitative methods in physical therapy research: a form of disciplined inquiry. *Phys Ther*;**69**:492–500.

Jick TD (1979). Mixing qualitative and quantitative methods: triangulation in action. *Admin Sci Q*;**24**:602–11.

Jirotka M, Goguen JA, editors (1994). Requirements engineering: social and technical issues. London; Academic Press.

Joas H (1985). GH Mead: a contemporary re-examination of his thought. Cambridge: Polity Press.

Johnson J (1990). Selecting ethnographic informants. In: Miller ML, Manning PK, Van Maanen J, editors. Qualitative research methods. Newbury Park, CA: Sage.

Jones K, Mullee M (1995). Pro-active, nurse-run asthma care in general practice: scientific fact or medical assumption? *Br J Gen Pract*;**45**:497–9.

Junker B (1952). Some suggestions for the design and field work learning experiences. In: Hughes E, editor. Cases on field work. Chicago: University of Chicago hectograph.

Jupp V, Norris C (1993). Traditions in documentary analysis. In: Hammersley M, editor. Social research: philosophy, politics and practice. London: Sage.

Kaplan B, Maxwell JA (1994). Qualitative research methods for evaluating computer information systems. In: Anderson JG, Aydin CE, Jay SJ, editors (1994). Evaluating health care informations: methods and applications. London: Sage. p. 45–68.

Kaplan B (1986). Impact of a clinical laboratory computer system: users' perceptions. In: Salmon R, Blum BI, Jorgensen MJ, editors. Medinfo 86: Fifth World Congress on Medical Infomatics. Amsterdam: North Holland. p. 1057–61.

Kaplan B (1987). Initial impact of a clinical laboratory computer system: themes common to expectations and actualities. *J Med Syst*;**11**:137–47.

Kaplan B, Duchon D (1988). Combining qualitative and quantitative methods in information systems research: a case study. *MIS Q*;**12**:571–86.

Kaplan B, Duchon D (1989). A job orientation model of impact on work seven months post-implementation. In: Barber B, Cao D, Qin D, Wagner G, editors. Medinfo 89: Sixth World Congress on Medical Infomatics. Amsterdam: North Holland. p. 1051–55.

Kant I (1964). Immanuel Kant's critique of Pure reason. London: Macmillan.

Kant I (1976).Critique of Practical reason, and other writings in moral philosophy. New York: Garland.

Katz J (1983). A theory of qualitative methodology: the social system of analytic fieldwork. In: Emerson RM, editor. Contemporary field research: a collection of readings. Prospect Heights, Ill: Waveland Press. p. 127–48.

Keen J, Packwood T (1995). Qualitative research: case study evaluation. *BMJ*;**311**:444–6.

Kelle U, editor (1995). Computer-aided qualitative data analysis. Theory, methods and practice. London: Sage.

Kelle U, Laurie H (1995). Computer use in qualitative research and issues of validity: In: Kelle U, editor. Computer-aided qualitative data analysis. London: Sage. p. 19–28.

Kelman H (1982). Ethical issues in different social science methods. In: Beauchamp T, Faden R, Wallace R, Walters L, editors. Ethical issues in social science research. Baltimore: Johns Hopkins University Press.

Kennedy M (1979). Generalizing from single case studies. *Evaluation Quarterly*;**3**:661–78.

Khan ME, Anker M, Patel BC, Barge S, Sadhwani H, Kohle R (1991). The use of focus groups in social and behavioural research: some methodological issues. *World Health Stat Q*;**44**(3):145–9.

Kinston W (1983). Pluralism in the organization of health services research. *Soc Sci Med*;**17**:299–313.

Kiresuk TJ, Larsen NE, Lund SH (1981). Management and evaluation in a knowledge transfer context. In: Levine RA, Solomon MA, Hellstern G, Wollman H, editors. Evaluation research and practice. Comparative and international perspectives. Beverley Hills, CA: Sage. p. 198–219.

Kirk J, Miller M (1986). Reliability and validity in qualitative research. Newbury Park, CA: Sage.

Kitzinger J (1995). Qualitative research: introducing focus groups. *BMJ*;**311**:299–302.

Kleinman S, Stenross B, McMahon M (1994). Privileging fieldwork over interviews: consequences for identity and practice. *Symbolic Interaction*;**17**:37–50.

Klockars C (1977). Field ethics for the life history. In: Weppner R, editors. Street ethnography. Beverley Hills, CA: Sage. p. 201–26.

Knafl KA, Breitmayer B (1991). Triangulation in qualitative research. In: Morse J, editor. Qualitative nursing research: a contemporary dialogue. Newbury Park, CA: Sage. p. 226–39.

Kolakowski L (1993). An overall view of positivism. In: Hammersley M, editor. Social research: philosophy, politics and practice. London: Sage. p. 1–8.

Krantz DL (1995). Sustaining vs. resolving the quantitative-qualitative debate. *Evaluation and Program Planning*;**18**:89–96.

Krefting L (1991). Rigor in qualitative research: the assessment of trustworthiness. *Am J Occup Ther*;**45**:214–22.

Krueger R (1994). Focus groups: a practical guide for applied research. Thousand Oaks, CA: Sage.

Kuhn T (1962). The structure of scientific revolutions. Chicago: University of Chicago Press.

Kuper A (1975). Anthropologist and anthropology: the British school 1922–72. Harmondsworth: Penguin.

Kuzel A (1986). Naturalistic inquiry: an appropriate model for family medicine. *Family Medicine*;**18**(6): 369–374.

Lacey C (1970). Hightown Grammar. Manchester: Manchester University Press.

Laing RD (1965). The divided self. Harmondsworth: Penguin.

Lasswell HD, Lerner D, Pool I (1952). The comparative study of symbols. Stanford CA: Stanford University Press.

Lateiner D.(1989). The historical method of Hirodotus. Toronto: University of Toronto Press.

Lather P (1986). Research as praxis. *Harvard Educ Rev*;**56**:257–77.

LeCompte M, Goetz JP (1982). Problems of reliability and validity in ethnographic research. *Rev Educ Res*;**52**:31–60.

LeCompte M, Preissle J (1993a). Ethnography and qualitative design in educational research. 2nd edition. New York: Academic Press Inc.

LeCompte M, Preissle J (1993b). Selecting and sampling in qualitative research. In: LeCompte M, Preissle J, editors. Ethnography and qualitative design in educational research. 2nd edition. London: Academic Press. p. 56–85.

LeCompte M, Preissle J, (1993c). Evaluating qualitative design. In: LeCompte M, Preissle J, editors. Ethnography and qualitative design in educational research. 2nd edition. London: Academic Press. p. 315–56.

LeCompte M, Preissle J (1993d). Selecting a research design. In: LeCompte M, Preissle J, editors. Ethnography and qualitative design in educational research. 2nd edition. London: Academic Press. p. 30–55.

Lee RM, Fielding NG (1991). Computing for qualitative research: options, problems and potential. In: Fielding NG, Lee RM, editors. Using computers in qualitative research. London: Sage. p. 1–13.

Lee RM, Fielding NG (1995). Users' experience of qualitative data analysis software. In Kelle U, editor. Computer-aided qualitative data analysis. Theories, methods and practice. London: Sage. p. 29–40.

Leese B, Bosanquet N (1995). Family doctors and change in practice strategy since 1986. *BMJ*;**310**:705–8.

Leininger M (1992). Current issues, problems and trends to advance qualitative paradigmatic research methods for the future. *Qualitative Health Research*;**12**:392–415.

Levine RA (1981). Program evaluation and policy analysis in Western nations. An overview. In: Levine RA, Solomon MA, Hellstern G, Wollman H, editors. Evaluation research and practice. Comparative and international perspectives. Beverley Hills, CA: Sage. p. 27–60.

Lincoln YS, Guba EG (1985). Naturalistic inquiry. Newbury Park, CA: Sage.

Lincoln YS (1990). The making of a constructivist: a remembrance of transformations past. In: Guba E, editor. The paradigm dialog. Newbury Park, CA: Sage. p. 67–87.

Lindesmith A (1947). Opiate addiction. Bloomington, Ind: Principia Press.

Lindesmith A, Strauss AL (1949). Social psychology. New York: Dryden Press.

Lindholm L, Ekbom T, Dash C, Eriksson M, Tibblin G, Schersten B, on behalf of the CELL study group (1995). The impact of health-care advice given in primary care on cardio-vascular risk. *BMJ*;**310**:1105–9.

Lindlof TR (1995). Qualitative communication research methods. London: Sage.

Lindsay J, O'Connell DC (1995). How do transcribers deal with audio recordings of spoken discourse? *J Psycholinguist Res*;**24**:101–15.

Lofland J, Lejeune R (1960). Initial interaction of newcomers in Alcoholics Anonymous: a field experiment in class symbols and socialization. *Social Problems*;**8**:102–11.

Lofland J (1967). Notes on naturalism. *Kans J Sociol*;**3**:45–61.

Lofland J (1971). Analyzing social settings: a guide to qualtitative observation and analysis. Belmont, CA: Wadsworth.

Lofland J (1976). Doing social life: the qualitative study of human interaction in natural settings. New York: Wiley.

Lonkila M (1995). Grounded Theory as an emerging paradigm for computer-assisted qualitative data analysis. In: Kelle U, editor. Computer-assisted qualitative data analysis, theories, methods and practice. London: Sage.

Lord J, Green F (1995). Exercise on prescription: does it work? *Health Educ J*;**54**:453–64.

Luff P, Heath CC, Greatbatch D (1994). Work, interaction and technology: the naturalistic analysis of human conduct and requirements capture. In: Jirotka M, Goguen J, editors. Requirements engineering: social and technical issues. London: Academic Press. p. 259–88.

Lindberg G (1993). Is soiciology too scientific? *Sociologus*;**9**:298–322.

Lundsgaarde HP, Fischer PJ, Steele DJ (1981). Human problems in realized medicine. University of Kansas Publications in Anthropology No. 13, Lawrence: University of Kansas.

Lyman S, Scott M (1989). A sociology of the absurd. 2nd edition. New York: General Hall, Inc.

Lynd RS, Lynd HM (1929). Middletown: a study in American culture. London: Constable.

Lynd RS, Lynd HM (1937). Middletown in transition: a study of cultural conflicts. London: Constable.

Lyytinen K (1987). Different perspectives on information systems: problems and solutions. *ACM Computing Surveys*;**19**:5–46.

Lyytinen K, Hirschheim R (1987). Information systems failure – a survey and classification of empirical literature. *Oxford Surveys in Information Technology*;**4**:257–309.

Macdonald K, Tipton K (1993). Using documents. In: Gilbert N, editor. Researching social life. London: Sage. p. 187–200.

MacIntyre A (1982). Risk, harm and benefits assessments as instruments of moral evaluation. In: Beauchamp T, Faden R, Wallace R, Walters L, editors. Ethical issues in social science research. Baltimore: Johns Hopkins. p. 175–89.

Macintyre SC (1977). Single and pregnant. London: Croom Helm.

Macklin R (1982). The problem of adequate disclosure in social science research. In: Beauchamp T, Faden R, Wallace R, Walters L, editors. Ethical issues in social science research. Baltimore: Johns Hopkins. p. 193–214.

Macnabb DGC (1966). David Hume: his theory of knowledge and morality. 2nd edition. Oxford: Basil Blackwell.

Maine H (1861). Ancient law. London: John Murray.

Malinowski B (1922). Argonauts of the Western Pacific. London: George Routledge & Sons.

Mangabeira W (1995). Qualitative analysis and microcomputer software. Some reflections on a new trend in sociological research. In: Burgess RG, editor. Computing and qualitative research. Greenwich, CA: JAI Press. p. 43–62.

Marsden D, Gissane C (1992). Nursing evaluation: purposes, achievements and opportunities. *Int J Nurs Stud*;**29**:231–36.

Marshall C (1985). Appropriate criteria of the trustworthiness and goodness for qualitative research on educational organizations. *Qual Quant*;**19**:353–73.

Marshall C, Rossman G (1989). Designing qualitative research. Newbury Park, CA: Sage.

Martineau H 1838). How to observe morals and manners. London: Charles Knight.

Mason J (1996). Qualitative researching. London: Sage.

Mass Observation (1938). First year's work 1937–1938. London: Lindsay Drummond.

Mass Observation (1949). Meet yourself at the doctor's. London: Naldrett Press.

Mathews C, Everett K, Binedell J, Steinberg M (1995). Learning to listen: formative research in the development of AIDS education for secondary school students. *Soc Sci Med*;**41**(12):1715–24.

Mathison S (1988). Why triangulate? *Educational Researcher*;**17**:13–17.

Matza D (1969). Becoming deviant. Englewood Cliffs, NJ: Prentice-Hall.

Maykut P, Morehouse R (1994). Beginning qualitative research: a philosophic and practical guide. London: Falmer Press.

Maynard D (1989). On the ethnography and analysis of discourse in institutional settings. *Perspectives on Social Problems*;**1**:127–46.

Maynard D, Clayman S (1991). The diversity of ethnomethodology. *Annu Rev Sociol*;**17**:385–418.

McCall G, Simmons J (1969). Issues in participant observation: a text and reader. Reading, MA: Addison-Wesley.

McCracken G (1988). The long interview. Newbury Park, CA: Sage.

McDonald L (1994). The women founders of the social science. Ottawa: Carleton University Press.

McHugh P (1970). A common-sense conception of deviance. In: Douglas J, editor. Deviance and respectability: the social construction of moral meanings. New York: Basic Books.

McKeganey N, Bloor M (1981). On the retrieval of sociological description: respondent validation and the critical case of ethnomethodology. *Int J Sociol Soc Policy*;**1**:58–69.

McKeganey N, Bloor M (1991). Spotting the invisible man: the influence of male gender on fieldwork relations. *Br J Sociol*;**42**:195–210.

McKenna H, Parahoo K, Boore J (1995a). The evaluation of a nursing model for long-stay psychiatric nursing care. Part 1 – Literature review and methodology. *Int J Nurs Stud*;**32**(1):79–94.

McKenna H, Parahoo K, Boore J (1995b). The evaluation of a nursing model for long-stay psychiatric nursing care. Part 2 – Presentation and discussion of findings. *Int J Nurs Stud*;**32**(1):95–113.

McKinlay TB (1993). The promotion of health through socio-political change: challenges for research and policy. *Soc Sci Med*;**36**:109.

Mechanic D (1978). Prospects and problems in health services research. *Milbank Memorial Fund Quarterly*;**56**:127–39.

Mechanic D (1989). Medical sociology: some tensions among theory, method and substance. *J Health Soc Behav*;**30**:147–60.

Medley DM (1963). Measuring classroom behaviour by systematic observation. In: Gage NL, editor. Handbook of research on teaching. Chicago: Round McNally.

Mehan J (1978). Structuring school structure. *Harvard Educational Review*;**48**:32–64.

Melia K (1996). Re-discovering Glaser. *Qual Health Res*;**6**:368–78.

Melia K (1997). Producing 'plausible stories' - interviewing student nurses. In: Miller G, Dingwall R, editors. Context and method in qualitative research. London: Sage. p. 26–36.

Merriam S (1988). Case study research in education: a qualitative approach. San Francisco, CA: Jossey Bass Inc.

Merton RK, Reader GC, Kendall PL (1957). The student physician: introductory studies in the sociology of medical education. Cambridge, MA: Harvard University Press.

Mies M (1983). Towards a methodology for feminist research. In: Bowles G, Duelli Klein R, editors. Theories of women's studies. London: Routledge & Kegan Paul.

Miles MB (1979). Qualitative data as an attractive nuisance: the problem of analysis. *Admin Sci Q*;**24**:590–601.

Miles MB, Huberman AM (1984). Qualitative data analysis. 1st edition. London: Sage.

Miles MB, Huberman MA (1994). Qualitative data analysis. An expanded sourcebook. 2nd edition. Thousand Oaks, CA: Sage.

Milgram S (1963). Behavioural study of obedience. *J Abnorm Soc Psychol*;**67**:371–8.

Mill JS (1973). A system of logic ratiocinative and inductive: being a connected view of the principles of evidence and the methods of scientific investigation: Books I–III and appendices. Toronto/London: University of Toronto Press/Routledge & Kegan Paul.

Mill JS (1974). A system of logic ratiocinative and inductive: being a connected view of the principles of evidence and the methods of scientific investigation: Books iv–vi and appendices. Toronto/London: University of Toronto Press/Routledge & Kegan Paul.

Miller G (1997). Context and method in qualitative methods. In: Miller G, Dingwall R, editors. Context and method in qualitative research. London: Sage.

Miller N, Morgan D (1993). Called to account: the CV as an autobiographical practice. *Sociology*;**27**(1):133–43.

Miller W, Crabtree B (1994). Clinical research. In: Denzin N, Lincoln Y, editors. Handbook of qualitative research. Thousand Oaks, CA: Sage.

Mishler EG (1979). Meaning in context: is there any other kind? *Harvard Educational Review*;**49**:1–19.

Mitchell ES (1986). Multiple triangulation: a methodology for nursing science. *JAdv Nurs*;**2**:261–70.

Mitchell JC (1983). Case and situational analysis. *Sociol Rev*;**31**:187–211.

Morgan D (1981). Men, masculinity and the process of sociological enquiry. In: Roberts H, editor. Doing feminist research. London: Routledge & Kegan Paul.

Morgan D (1993). Successful focus groups. Newbury Park, CA: Sage.

Morgan D (1996). Focus groups: a new tool for qualitative research. *Qual Sociol*;**7**:253–70.

Morgan G (1983). Beyond method. Beverley Hills, CA: Sage.

Morgan M, Watkins C (1988). Managing hypertension: beliefs and responses to medication among cultural groups. *Sociology of Health and Illness*;**10**:561–78.

Morrison A, McIntyre D (1969). Teachers and teaching. Harmondsworth: Penguin.

Morse J (1994). Designing funded qualitative research. In: Denzin N, Lincoln Y, editors. Handbook of qualitative research. Thousand Oaks, CA: Sage.

Mumford E, Hirschheim R, Fitzgerald G, Wood-Harper T, editors (1985). Research methods in information systems. Amsterdam: North Holland.

Munhall P (1993a). Ethical considerations in qualitative research. In: Munhall PL, Oiler Boyd C, editors. Nursing research: a qualitative perspective. 2nd edition. New York: National League for Nursing. p. 395–408.

Munhall P (1993b). Language and nursing research. In: Munhall PL, Oiler Boyd C, editors. Nursing research: a qualitative perspective. New York: National League for Nursing Research. p. 5–38.

Munhall P (1993c). Epistemology in nursing. In: Munhall PL, Oiler Boyd C, editors. Nursing research: a qualitative perspective. 2nd edition. New York: National League for Nursing. p. 39–65.

Nader L (1969). Up the anthropologist: perspectives gained from studying up. In: Hymes D, editor. Reinventing anthropology. New York: Vintage. p. 284–311.

Nagel E, editor (1950). John Stuart Mill's philosophy of scientific method. New York: Hafner Publishing Company.

Newens A, McEwan R (1995). AIDS/HIV awareness training for young people with severe learning difficulties: an evaluation of two school programmes. *JAdv Nurs*;**22**:267–75.

Nissen HE, Klein HK, Hirschheim R, editors (1991). Information systems research: contemporary approaches and emergent traditions. Amsterdam: North Holland.

Nolan M, Behi R (1995). Triangulation: the best of all worlds? *Br J Nurs*;**4**:829–32.

Nyce JM, Graves W III (1990). The construction of neurology: implications for hypermedia system development. *Artificial Development in Medicine*;**2**:315–22.

Oakley A (1981). Interviewing women: a contradiction in terms. In: Roberts H, editor. Doing feminist research. London: Routledge & Kegan Paul. p. 30–61.

Oakley A (1984). Taking it like a woman. London: Cape.

Oiler Boyd C (1993a). Combining qualitative and quantitative approaches. In: Munhall PL, Oiler Boyd C, editors. Nursing research: a qualitative perspective. 2nd edition. New York: National League for Nursing Press. p. 454–75.

Oiler Boyd C (1993b). Philosophical foundations of qualitative research. In: Munhall PL, Oiler Boyd C, editors. Nursing research: a qualitative perspective. 2nd edition. New York: National League for Nursing Press. p. 66–93.

Olesen VL (1994). Feminisms and models of qualitative research. In: Denzin NK, Lincoln YS, editors. Handbook of qualitative research. Thousand Oaks, CA: Sage. p. 158–74.

Ong BN (1993). The practice of health services research. London: Chapman & Hall.

Ormala E (1994). Impact assessment. European experience of qualitative methods and practices. *Evaluation Review*;**18**:41–51.

Owens R (1982). Methodological rigour in naturalistic inquiry. *Educ Admin Q;***18**:1–21.

Paget MA (1983). Experience and knowledge. *Hum Stud*;**6**:67–90.

Palmer VM (1928). Field studies in sociology: a student's manual. Chicago: University of Chicago Press.

Parlett M, Hamilton D (1977). Evaluation as illumination; a new approach to the study of innovatory programmes. In: Hamilton D, Jenkins D, King C, MacDonald B, Parlett M, editors. Beyond the numbers game. Basingstoke: Macmillan Education. p. 3–22.

Parry O, Pill R (1994). I don't tell him how to live his life: the doctor/patient encounter as an educational context. In: Bloor M, Taraborelli P, editors. Qualitative studies in health and medicine. Aldershot: Avebury. p. 5–21.

Parsons T (1951). The social system. London: Routledge & Kegan Paul.

Paterson C, Peacock W (1995). Complementary practitioners as part of the primary health care team: evaluation of one model. *Br J Gen Pract;* **45**:255–8.

Patterson BR, Neupauer NC, Burant PA, Koehn SC, Reed AT (1996). A preliminary examination of conversation analytic techniques – rates of inter-transcriber reliability. *West J Commun*;**60**:76–91.

Patton MQ (1980). Qualitative evaluation methods. Newbury Park, CA: Sage.

Patton MQ (1986). Utilization-focused evaluation. 2nd edition. Newbury Park, CA: Sage.

Patton MQ (1990). Qualitative education and research methods. 2nd edition. Newbury Park, CA: Sage.

Pattullo E (1982). Modesty is the best policy: the federal role in social research. In: Beauchamp T, Faden R, Wallace R, Walters L, editors. Ethical issues in social science research. Baltimore: Johns Hopkins. p. 373–90.

Pearsol J (1985). The nature of explanation in qualitative education. *Evaluation and the Health Professions*;**8**:129–47.

Pettinari CJ (1988). Task, talk and text in the operating room: a study in medical discourse. Norwood, NJ: Ablex.

Pfaffenberger B (1988). Microcomputer applications in qualitative research. Beverley Hills, CA: Sage.

Phillips D (1987). Validity in qualitative research: why the worry about warrant will not wane. *Education and Urban Society*;**20**:9–24.

Phillips D (1990). Post-positivistic science: myths and realities. In: Guba E, editor. The paradigm dialogue. Newbury Park, CA: Sage. p. 31–45.

Pill R (1995). Fitting the method to the question: the quantitative or qualitative approach? In: Jones R, Kinmonth A, editors. Critical reading for primary care. Oxford: Oxford University Press. p. 42–58.

Pollard A (1995). Opportunities and difficulties of a teacher-ethnographer: a personal account. In: Burgess RG, editor. Field methods in the study of education. Lewes: Falmer.

Polsky N (1971). Hustlers, beats and others. Harmondsworth: Penguin.

Popper K (1959). The logic of scientific discovery. London: Hutchinson.

Platt J (1994). The Chicago School and firsthand data. *History of the Human Sciences*;**7**:57–80.

Platt J (1996). History of sociological research methods in America 1920–1960. Cambridge: Cambridge University Press.

Prasad B, Costello A (1995). Impact and sustainability of a 'baby friendly' health education intervention at a district hospital in Bihar, India. *BMJ*;**310**:621–3.

Prior L (1985). Making sense of mortality. *Sociology of Health and Illness*;**7**(2):167–90.

Prout A (1992). Illumination, collaboration, facilitation, negotiation: Evaluating the MESMAC project. In: Aggleton P, Young A, Moody D, Kapila M, Pye M, editors. Does it work? Perspectives on the evaluation of HIV/AIDS health promotion. London: Health Education Authority. p. 77–91.

Psathas G (1968). The student nurse in the Diploma School of Nursing. New York: Springer Publishing Company. p. 79–118.

Punch M (1994). Politics and ethics in qualitative research. In: Denzin N, Lincoln Y, editors. Handbook of qualitative research. Thousand Oaks, CA: Sage. p. 83–97.

Ragin CC (1994). Constructing social research. The unity and diversity of method. Thousand Oaks, CA: Pine Forge Press.

Randall D, Hughes J, Shapiro D (1994). Steps towards a partnership: ethnography and system design. In: Jirotka M, Goguen J, editors. Requirements engineering: social and technical issues. London: Academic Press. p. 241–58.

Rees C (1981). Records and hospital routine. In: Atkinson P, Heath C, editors (1981). Medical work: realities and routines. Aldershot: Gower.

Reichardt CS, Cook TD (1979). Beyond qualitative versus quantitative methods. In: Cook TD, Reichardt CS, editors. Qualitative and quantitative methods in evaluation research. Beverley Hills, CA: Sage. p. 7–32.

Reichenbach H (1938). Experience and prediction: an analysis of the foundations and the structure of knowledge. Chicago: University of Chicago Press.

Reichenbach H (1951). The rise of scientific philosophy. Berkeley, CA. University of California Press.

Reinharz S (1994). Qualitative evaluation and policy. In: Gubrium JF, Sankar A, editors. Qualitative methods in aging research. Thousand Oaks, CA: Sage. p. 263–76.

Reiss AJ (1979). Governmental regulation of scientific inquiry: some paradoxical consequences. In: Klockars C, O'Connor F, editors. Deviance and decency: the ethics of research with human subjects. Beverley Hills, CA: Sage. p. 61–95.

Rich RF (1981). Can evaluation be used for policy action? In: Levine RA, Solomon MA, Hellstern G, Wollman H, editors. Evaluation research and practice. Comparative and international perspectives. Beverley Hills, CA: Sage. p. 178–91.

Richards L, Richards T (1991). The transformation of qualitative method: computational paradigms and research processes. In: Fielding NG, Lee RM, editors. Using computers in qualitative research. London: Sage. p. 38–53.

Richards TJ, Richards L (1994a). From filing cabinet to computer. In: Bryman A, Burgess RG, editors. Analysing qualitative data. London: Routledge. p. 146–72.

Richards TJ, Richards L (1994b). Using computers in qualitative research. In: Denzin N, Lincoln Y, editors. Handbook of qualitative research. Thousand Oaks, CA: Sage. p. 445–62.

Richards L (1995). Transition work! Reflections on a three-year NUD.IST project. In: Burgess RG, editor. Computing and qualitative research. Greenwich CT: JAI Press. p. 105–40.

Richardson L (1994). Writing: a method of inquiry. In: Denzin NK, Lincoln YS, editors. Handbook of qualitative research. Thousand Oaks, CA: Sage. p. 516–29.

Richardson S, Dohrenwend B, Klein D (1965). Interviewing: its forms and functions. New York: Basic Books.

Rist RC (1977). On the relations among educational research paradigms: from disdain to detente. *Anthropol Educ Q*;**8**:42–9.

Rist RC (1995). Introduction. In: Rist RC, editor. Policy evaluation. Linking theory to practice. Aldershot: Elgar. xiii–xxvi.

Robinson D (1971). The process of becoming ill. London: Routledge & Kegan Paul.

Robinson J (1982). An evaluation of health visiting. London: CETHV.

Robinson WS (1951). The logical structure of analytic induction. *Am Sociolog Rev*;**16**:812–18.

Rock P (1979). The making of symbolic interactionism. London: Macmillan.

Roethlisberger FJ, Dickson WJ (1939). Management and the worker: an account of a research program conducted by the Western Electric Company, Hawthorne Works, Chicago. Cambridge, Mass: Harvard University Press.

Rohner RP (1966). Franz Boas: ethnographer on the Northwest Coast. In: Helm J, editor. Pioneers of American anthropology: the uses of biography. Seattle: University of Washington Press.

Roman L, Apple M (1990). Is naturalism a move away from positivism? Materialist and feminist approaches to subjectivity in ethnographic research. In: Eisner E, Peshkin A, editors. Qualitative inquiry in education: the continuing debate. New York: Teachers College Press. p. 38–73.

Rosenhan D (1973). On being sane in insane places. *Science*;**179**:250–8.

Rossi P (1981). The professionalization of evaluation research in the United States. In: Levine RA, Solomon MA, Hellstern G, Wollman H, editors. Evaluation research and practice. Comparative and international perspectives. Beverley Hills, CA: Sage. p. 220–36.

Rossi P, Freeman H (1993). Evaluation: a systematic approach. 5th edition. Newbury Park, CA: Sage.

Roth J (1963). Timetables: structuring the passage of time in hospital treatment and other careers. Indianapolis: Bobbs-Merill.

Roth J (1970). Comments on 'secret observation'. In: Filstead W, editor. Qualitative methodology: firsthand involvement with the social world. Chicago: Markham. p. 278–80.

Russell I (1996). Methods of health services evaluation: the gospel of Archie Cochrane after 25 years. *J Health Serv Res Policy*;**1**:114–15.

Ryan A (1970). The philosophy of John Stuart Mill. London: Macmillan.

Ryan A (1974). J.S. Mill. London: Routledge & Kegan Paul.

Sacks H, Schegloff EA, Jefferson J (1974). A simplest systematics for the organisation of turn-taking for conversation. *Language*;**50**:696–735.

St Leger AS, Schneiden H, Walsworth-Bell JP (1992). Evaluating health services' effectiveness. A guide for health professionals, service managers and policy managers. Milton Keynes: Open University Press.

Sandelowski M (1986). The problem of rigor in qualitative research. *Ans-Advances in Nursing Science*;**8**:27–37.

Sandelowski M (1994). Focus on qualitative methods: notes on transcription. *Res Nurs Health*;**17**:311–14.

Schatzman L, Strauss A (1973). Field research: strategies for a natural sociology. Englewood Cliffs, NJ: Prentice-Hall.

Scheff TJ (1963). The role of the mentally ill and the dynamics of mental disorder: a research framework. *Sociometry*;**26**:436–53.

Schegloff EA (1997). Whose text? Whose context? *Discourse and Society*;**8**(2):165–87.

Schegloff EA, Sacks H (1973). Opening up closings. *Semiotica*;**8**:289–327.

Schneider L, editor (1967). The Scottish moralists on human nature and society. Chicago: University of Chicago Press.

Schofield J (1993). Increasing the generalizability of qualitative research. In: Hammersley M, editor. Social research: philosophy, politics and practice. London: Sage. p. 200–25.

Schutz A (1962). Collected papers, Vol 1: the problem of social reality. The Hague: Martinus Nijhoff.

Schutz A (1967). The phenomenology of the social world. Evanston, Ill: Northwestern University Press.

Schwandt T (1997). Qualitative inquiry: a dictionary of terms. Thousand Oaks, CA: Sage.

Scott M, Lyman S (1968). Accounts. *Am Sociolog Rev*;**33**:46–62.

Scott M, Lyman S (1970). Accounts, deviance and social order. In: Douglas J, editor. Deviance and respectability. New York: Basic Books. p. 89–119.

Scott S (1984). The personable and the powerful: gender and status in sociological research. In: Bell C, Roberts H, editors. Social researching: politics, problems and practice. London: Routledge & Kegan Paul. p. 165–78.

Scott S (1992). Evaluation may change your life but it won't solve all your problems. In: Aggleton P, Young A, Moody D, Kapila M, Pye M, editors. Does it work? Perspectives on the evaluation of HIV/AIDS health promotion. London: Health Education Authority. p. 57–75.

Scriven M (1991). Evaluation thesaurus. 4th edition. Newbury Park, CA: Sage.

Scruton R (1982). Kant. Oxford: Oxford University Press.

Secker J, Wimbush E, Watson J, Milburn K (1995). Qualitative methods in health promotion research: some criteria for quality. *Health Educ J*;**54**:74–87.

Seidel J (1991). Method and madness in the application of computer technology to qualitative data analysis. In: Fielding NG, Lee RM, editors. Using computers in qualitative data. London: Sage. p. 107–16.

Seidel J, Kelle U (1995). Different functions of coding in the analysis of textual data. In: Kelle U, editor. Computer-aided qualitative data analysis, theory, methods and practice. London: Sage. p. 29–40.

Sellitz C, Jahoda M, Deutsch M, Cook S (1964). Research methods in social relations. New York: Holt, Rinehart & Winston.

Shadish WRJ, Epstein R (1987). Patterns of program evaluation practice among members of the Evaluation Research Society and Evaluation Network. *Evaluation Review*;**11**:555–90.

Shadish WR, Cook TD, Leviton LC (1991). Foundations of program evaluation: theories of practice. Newbury Park, CA: Sage.

Shadish WR (1995). The quantitative-qualitative debates: 'DeKuhnifying' the conceptual context. *Evaluation and Program Planning*;**18**:47–9.

Shaffir W, Stebbins R, Turowetz A (1980). Fieldwork experience: qualitative approaches to social research. New York: St Martin's Press.

Shaffir W (1985). Some reflections on approaches to fieldwork in Hassidic communities. *Jewish Journal of Sociology*;**27**:115–34.

Shaw M, Miles I (1979). The social roots of statistical knowledge. In: Irvine J, Miles I, Evans J, editors. Demystifying social statistics. London: Pluto Press. p. 27–38.

Shiroyama C, McKee L, McKie L (1995). Evaluating health promotion projects in primary care: recent experiences in Scotland. *Health Educ J*;**54**:226–40.

Shneiderman B, Carroll JM (1988). Ecological studies of professional programmers: an overview. *Communications of the ACM*;**31**:1256–8.

Shore M (1987). The science of social redemption: McGill, the Chicago School and the origins of social research in Canada. Toronto: University of Toronto Press.

Shortell SM, Richardson WC (1978). Health program evaluation. St Louis: CV Mosby.

Silverman D (1981). The child as a social object: Down's syndrome children in a paediatric cardiology clinic. *Sociology of Health and Illness*;**3**:254–74.

Silverman D (1984). Going private: ceremonial forms in a private oncology clinic. *Sociology*;**18**:191–204.

Silverman D (1985). Qualitative methodology and sociology. 1st edition, Aldershot: Gower.

Silverman D (1989). Telling convincing stories: a plea for cautious positivism in case studies. In: Glassner B, Moreno JD, editors. The qualitative-quantitative distinction in the social sciences. Dordrecht: Kluwer Academic Publishers. p. 57–77.

Silverman D, Bor R, Miller R, Goldman E (1992). Obviously the advice is then to keep to safer sex: Advice giving and advice reception in AIDS counselling. In: Aggleton P, Davies P, Hart G, editors. AIDS: rights, risks and reason. London: Falmer. p. 174–91.

Silverman D (1993). Interpreting qualitative data: methods for analysing talk, text and interaction. London: Sage.

Silverman D (1997). The logics of qualitative research. In: Miller G, Dingwall R, editors. Context and method in qualitative research. London: Sage.

Skrtic T, Guba EG, Knowlton H (1985). Special education in rural America: interorganizational special education programming in rural areas. Washington DC: Department of Education, National Institute of Education.

Small AK, Vincent GE 1894). Introduction to the science of sociology. New York: American Book Company.

Smith A (1976). An inquiry into the nature and causes of the wealth of nations. Chicago: University of Chicago Press.

Smith A (1976). The theory of moral sentiments. Oxford: Clarendon Press.

Smith D, Pearce L, Pringle M, Caplan R (1995). Adults with a history of child sexual abuse: evaluation of a pilot therapy service. *BMJ*;**310**:1175–8.

Smith G, Cantley C (1985). Assessing health care: a study in organizational evaluation. Milton Keynes: Open University Press.

Smith H (1975). Strategies of social research: the methodological imagination. London: Prentice Hall.

Smith J (1983a). Quantitative versus interpretive: the problem of conducting social inquiry. In: House E, editor. Philosophy of evaluation. San Francisco: Jossey Bass. p. 27–51.

Smith J (1983b). Quantitative versus qualitative research: an attempt to clarify the issue. *Educational Researcher*;**12**:6–13.

Smith J (1984). The problem of criteria for judging interpretive inquiry. *Educational Evaluation and Policy Analysis*;**6**:379–91.

Smith J (1985). Social reality as mind-dependent versus mind-independent and the interpretation of test validity. *J Res Dev Educ*;**19**(1):1–9.

Smith J, Heshusius L (1986). Closing down the conversation: the end of the quantitative-qualitative debate among educational inquirers. *Educational Researcher*;**15**:4–12.

Smith J (1989). The nature of social and educational inquiry. Norwood, NJ: Ablex.

Smith R (1995). The scientific basis of health services. Needs a better balance between original and evaluative research. (Editorial). *BMJ*;**311**:961–2.

Spiers R, Jewell J (1995). One counsellor, two practices: report of a pilot scheme in Cambridgeshire. *Br J Gen Pract*;**45**:31–3.

Spencer H (1874). Descriptive sociology. London: Williams & Norgate.

Spencer H (1876). Principles of sociology. London: Williams & Norgate.

Stacey M (1960). Tradition and change: a study of Banbury. Oxford: Oxford University Press.

Stacey M, Dearden R, Pill R, Robinson D (1970). Hospitals, children and their families. London: Routledge & Kegan Paul.

Stacey M, Batstone E, Bell C, Murcott A (1975). Power, persistence and change: a second study of Banbury. London: Routledge & Kegan Paul.

Stake R (1986). Quieting reform. Urbana, Ill.: University of Illinois Press.

Stake R (1994). Case studies. In: Denzin N, Lincoln Y, editors. Handbook of qualitative research. Thousand Oaks, CA: Sage. p. 236–47.

Stake R (1995). The art of case study research. Thousand Oaks, CA: Sage.

Stange K, Zyzanski S (1989). Integrating qualitative and quantitative research methods. *Family Medicine*;**21**:448–51.

Stanley L, Wise S (1983). Breaking out: feminist consciousness and feminist research. London: Routledge & Kegan Paul.

Stanley L (1992). On auto/biography in sociology. *Sociology*;**27**(1):41–52.

Stanley L, Temple B (1995). Doing the business: using qualitative software packages in the analysis of qualitative datasets. In: Burgess RG, editor. Computing and qualitative research. Greenwich, CT: JAI Press.

Steckler A (1989). The use of qualitative evaluation methods to test internal validity: an example in a work site health promotion program. *Evaluation and the Health Professions*;**12**:115–33.

Steckler A, McLeroy KR, Goodman RM, Bird ST, McCormick L (1992). Towards integrating qualitative and quantitative methods: an introduction. *Health Educ Q*;**19**:1–8.

Stewart-Brown S, Surrender R, Bradlow J, Coulter A, Doll H (1995). The effects of fundholding in general practice on prescribing habits three years after introduction of the scheme. *BMJ*;**311**:1543–7.

Stimson G, Webb B (1975). Going to see the doctor. London: Routledge.

Stimson GV (1977). Do drug advertisements provide therapeutic information? *J Med Ethics*;**3**:7–13.

Stocking GW (1995). After Tylor: British social anthropology 1888–1951. Madison, WI: University of Wisconsin Press.

Strauss A, Schatzman L, Bucher L, Ehrlick D, Sabshin M (1964). Psychiatric ideologies and institutions. New York: Free Press.

Strauss A (1987). Qualitative analysis for social scientists. Cambridge: Cambridge University Press.

Strauss A, Corbin J (1990). Basics of qualitative research: grounded theory procedures and techniques. Newbury Park, CA: Sage.

Strong P (1979a). The ceremonial order of the clinic. London: Routledge & Kegan Paul.

Strong P (1979b). Sociological imperialism and the profession of medicine. *Soc Sci Med;***13**:199–215.

Strong P (1988). Qualitative sociology in the UK. *Qual Sociol*;**11**:13–28.

Suchman LA (1987). Plans and situated actions: the problem of human-machine interaction. Cambridge: Cambridge University Press.

Szasz TS (1972). The myth of mental illness. London: Paladin.

Taylor I, Walton P, Young J (1973). The new criminology: for a social theory of deviance. London: Routledge & Kegan Paul.

Tesch R (1990). Qualitative research: analysis types and software tools. London: Falmer.

Tesch R (1991). Software for qualitative researchers: analysis needs and program capabilities. In: Fielding NG, Lee RM, editors. Using computers in qualitative research. London: Sage. p. 16–37.

Thomas W, Thomas D (1927). The child in America: behaviour problems and programs. New York: Alfred Knopf.

Thomas W, Znaniecki F (1927). The Polish peasant in Europe and America. New York: Knopf.

Thomas W (1967). The unadjusted Girl. New York: Harper & Row.

Thorne B (1980). 'You still takin' notes?' Fieldwork and problems of informed consent. *Social Problems*;**27**:284–97.

Trend M (1979). On the reconciliation of qualitative and quantitative analyses. In: Reichardt C, Cook T, editors. Qualitative and quantitative methods in evaluation research. Beverley Hills, CA: Sage. p. 68–86.

Trow M (1969). Comment on 'participant observation and interviewing: a comparison'. In: McCall G, Simmons J, editors. Issues in participant observation: a text and reader. Reading, Mass: Addison-Wesley. p. 332–8.

Tuck R (1993). Philosophy and government 1572–1651. Cambridge: Cambridge University Press.

Tudor-Smith C, Frankland J, Playle R, Moore L (1995). Life education centres: an evaluation of a mobile health educational resource in Wales for children. *Health Educ J*;**54**:393–404.

Turner R (1969). The quest for universals in qualitative research. In: McCall G, Simmons J, editors. Issues in participant observation: a text and reader. Reading, Mass: Addison-Wesley. p. 205–16.

Twinn S, Shiu A (1995). An evaluation of the effectiveness of public health nursing: a case study of the maternal and child health centres in Hong Kong. *Int J Nurs Stud*;**33**(4):442–54.

Urry J (1993). Before social anthropology: essays on the history of British anthropology. Chur: Harwood Academic.

van der Walt H, Mathews C (1995). How do health service managers respond to qualitative research? *Soc Sci Med*;**41**:1725–9.

van Teijlingen E, Friend J, Twine F (1995). Evaluating clinical educators: an employer's perspective. *J Adv Nurs*;**21**:737–42.

Vidal De La Balche P (1931). Principles of human geography. New York: Henry Holt & Company.

Vidich A, Shapiro G (1969). A comparison of participant observation and survey data. In: McCall G, Simmons J, editors. Issues in participant observation: a text and reader. Reading, Mass: Addison-Wesley. p. 295–303.

Voysey M (1975). A constant burden: the reconstitution of family life. London: Routledge & Kegan Paul.

Waitzkin H (1979). Medicine, superstructure and politics. *Soc Sci Med*;**13A**:601–9.

Waitzkin H (1990). On studying the discourse of medical encounters. A critique of quantitative and qualitative methods and a proposal for a reasonable compromize. *Med Care*;**28**:473–88.

Walker M (1989). Analysing qualitative data: ethnography and the evaluation of medical education. *Med Educ*;**23**:498–503.

Walker R (1980). The conduct of educational case studies: ethics, theory and procedures. In: Dockerell W, Hamilton D, editors. Rethinking educational research. London: Hodder & Stoughton.

Walker R (1985). An introduction to applied qualitative research. In: Walker R, editor. Applied qualitative research. Aldershot: Gower. p. 3–26.

Walker R (1993). Evaluating drug education programmes. In: Colquhoun D, Kellehear A, editors. Health research in practice. Political, ethical and methodological issues. London: Chapman & Hall. p. 138–51.

Wallace PG, Haines AP (1984). General practitioners and health promotion: what patients think. *BMJ*;**289**:534–6.

Wallace W (1978). An overview of elements in the scientific process. In: Bynner J, Stribley K, editors. Social research: principles and procedures. London: Longman, in association with Open University Press. p. 4–10.

Walsham G (1993). Interpreting information systems in organizations. Chichester: Wiley.

Warner WL (1963). Yankee city. New Haven: Yale University Press.

Warren C (1988). Gender issues in field research. Newbury Park, CA: Sage.

Warren C, Rasmussen P (1977). Sex and gender in fieldwork research. *Urban Life*;**6**:359–69.

Warwick D (1982a). Tearoom trade: means and ends in social research. In: Bulmer M, editor. Social research ethics. London: Macmillan.

Warwick D (1982b). Types of harm in social research. In: Beauchamp T, Faden R, Wallace R, Walters L, editors. Ethical issues in social science research. Baltimore: Johns Hopkins. p. 101–24.

Wax M (1977). On fieldworkers and those exposed to fieldwork: federal regulations and moral issues. *Human Organization*;**36**:321–8.

Wax M, Cassell J (1979). Fieldwork, ethics and politics: the wider context. In: Wax M, Cassell J, editors. Federal regulations: ethical issues and social research. Boulder, Colorado: Westview. p. 85–102.

Wax M (1980). Paradoxes of 'consent' to the practice of fieldwork. *Social Problems*;**27**:272–83.

Wax R (1971). Doing fieldwork: warnings and advice. Chicago: University of Chicago Press.

Weaver A, Atkinson P (1994). Microcomputing and qualitative data analysis. Aldershot: Avebury.

Webb E, Campbell D, Schwartz R, Sechrest L, Grove J (1981). Non-reactive measures in the social sciences. 2nd edition. Boston: Hougton Mifflin.

Webb RK (1960). Harriet Martineau: a radical Victoria. London: Heinemann.

Weber M (1963). 'Objectivity' in social science and social policy. In: Natanson M, editor. Philosophy of the social sciences: a reader. New York: Random House.

Weick K (1968). Systematic observation. In: Lindzey G, Aronson E, editors. Handbook of social psychology. Reading, Mass: Addison-Wesley. p. 357–451.

Weiss C (1984). Toward the future of stakeholder approaches in evaluation. In: Conner R, Altman D, Jackson C, editors. Evaluation studies review annual volume 9. Beverley Hills, CA: Sage. p. 255–68.

Weitzman EA, Miles MB (1995). Computer programs for qualitative data analysts. Thousand Oaks, CA: Sage.

Wellard S, Rolls C, Ferguson de S (1995). Evaluating clinical educators: an employer's perspective. *J Adv Nurs*;**21**:737–42.

Werner O, Schoepfle G (1987). Systematic fieldwork, (2 vols). Newbury Park, CA: Sage.

West P (1990). The status and validity of accounts obtained at interview: a contrast between two studies of families with a disabled child. *Soc Sci Med*;**30**:1229–39.

Whalen J (1995). A technology of order production: computer-aided dispatch in public safety communications. In: ten Have P, Psthas G, editors. Situated order: studies in the social organization of talk and embodied activities. Washington, DC: International Institute for Ethnomethodology and Conversation Analysis. p. 185–230.

Whalen J (1996). Expert systems versus systems for experts: computer-aided dispatch as a support system in real world environments. In: Thomas PJ, editor. Social and interactional dimensions of human–computer interfaces. Cambridge: Cambridge University Press.

Whiteley S (1992). Evaluation of nursing education programmes – theory and practice. *Int J Nurs Stud*;**29**:315–23.

Wieder D (1985). Behavioristic operationalism and the life world: chimpanzees and chimpanzee researchers in face to face interaction. *Sociological Inquiry*;**50**:75–103.

Williams M, May T (1996). Introduction to the philosophy of social research. London: University College London.

Wilton J (1995). Preregistration house officers in general practice. *BMJ*;**310**:369–72.

Wilson S (1977). The use of ethnographic techniques in educational research. *Rev Educ Res*;**47**:245–65.

Wiseman J (1970). Stations of the lost: the treatment of skid row alcoholics. Englewood Cliffs, NJ: Prentice-Hall.

Wolcott HF (1975). Criteria for an ethnographic approach to research in schools. *Human Organization*;**34**:111–27.

Wolcott HF (1994). Transforming qualitative data. Description, analysis and interpretation. Thousand Oaks, CA: Sage.

Woodward CA (1993). Survivor focus groups: a quality assurance technique. *Palliative Medicine*;**7**(3):229–34.

World Health Organization (1981). Health programme evaluation. Guiding principles. Geneva: World Health Organization.

Young M, Wilmott P (1962). Family and kinship in East London. Harmondsworth: Penguin.

Young MDF, editor (1971). Knowledge and control: new directions for the sociology of education. London: Collier Macmillan.

Zelditch M (1969). Some methodological problems of field studies. In: McCall G, Simmons J, editor. Issues in participant observation: a text and reader. Reading, Mass: Addison-Wesley. p. 5–19.

Zerubavel E (1979). Patterns of time in hospital life. Chicago: University of Chicago Press.

Zimmerman DH (1970). The practicalities of rule use. In: Douglas JD, editor. Understanding everyday life. Chicago: Aldine. p. 221–38.

Zimmerman D (1988). On conversation: the conversation analytic perspective. In: Anderson JA, editor. Communication yearbook 11. Newbury Park, CA: Sage. p. 406–32.

Znaniecki F (1963). Cultural sciences: their origin and development. Urbana, Ill: University of Illinois Press.

Zubroff S (1988). In the age of the smart machine. New York: Basic Books.

Appendix 1

Summaries of selected studies discussed in the report

Baruch G (1981). Moral tales: parents' stories of encounters with the health profession. ***Sociology of Health and Illness*****;3:275–96.**

Baruch's research considered what parents were doing (or trying to do) when they related 'atrocity' stories of encounters with medical professionals in relation to their children's health. The central feature to all of these stories was an encounter with a medical professional where something went wrong. The stories are not analysed as an indicator of 'what really happened' in the encounters but were taken to be an account produced (or more likely, reproduced) for specific purposes within the interview setting. The atrocity stories were constructed in ways which allow them to be read as demonstrations of adequate parenthood.

In these atrocity stories, respondents presented two worlds (or realities): the parents' inhabited world, which was seen as that of the everyday or common sense; and the medical profession, which was presented as inhabiting a separate world for which additional knowledge and skills are equired and which, while respected, does not sit easily alongside the parental world. The parents acknowledged their inexperience in dealing with medical aspects of their child's welfare but maintained that health professionals expected them to be accomplished in the field. Parents could thus explain their inability to fulfil these expectations and at the same time display a status of adequate parenthood.

This study consisted of a series of semi-structured interviews with the parents of two groups of children. The children were either attending a paediatric cardiology unit or they were being treated for a cleft palate or hare lip at a suburban children's hospital. Thirty-two families were involved in the study. There were more respondents from the group of parents of children with coronary heart disease; in these cases it was possible to monitor the child's medical career. For the children with a cleft palate or hare lip a cross-sectional approach was adopted and a sample of children was selected from three age groups (0–3 years, 4–10 years, 11 years and over).

Interviews were conducted either in the respondents' homes or at the specialist hospitals and were tape-recorded. The interviews were divided into two parts. The first part of the interview consisted of a rather formal recounting of parents' experiences in a question and answer format, while the second part was more conversational in nature. The interviewer asked the respondents to tell the 'story of the child's career' and then to elaborate on the following issues:

- their understanding of the child's condition and its future prospects
- their reaction to the specialist hospital and to health professionals
- their versions of the cause of illness
- their capacity to cope at home
- arrangements that had been made for the child's hospital admission
- the reaction of any other siblings.

Baruch resisted the temptation to treat parents' atrocity stories as a more or less accurate representation of the encounters between parents and health professionals. Rather he treated the stories 'situated accounts' for the purpose of analysis. In other words he focused upon what the parents could be seen to be **doing** as they told these stories. The analysis aimed to examine the ways actors invoke structures rather than imposing further structure on the data. The stories were treated as displays of adequate parenthood. Six elements were identified as making up a framework in which the status of moral adequacy was established.

- The story tellers located themselves in a shared world.
- They located health professionals in a world distinct from that of the lay population.
- They appealed to features of their world to show that according to these they had acted reasonably, given the circumstances.
- Members of the medical world were often shown to have acted incompetently according to the standards of their world.
- The rules of the medical world were acknowledged as reasonable except when put into the context of the everyday world.

- By attending to the relevance of the medical standards and rules, they enhanced their status as competent members.

Four features were identified from within the framework by which parents constructed the two realities and displayed their status of moral adequacy. These were emotionality, child rearing practices, priorities in outpatient clinics and the meanings of diagnostic labels.

Baruch ended with a brief discussion of how this framework could be applied practically by health professionals to allay parental fears early on in the child's medical career. He argued that only through a proper analysis of the presentation of atrocity stories (rather than taking them at face value or appealing to external structures for analysis) will a balanced interpretation be possible.

This study illustrates the dangers of treating respondents' interview accounts as reproductions of the events they describe. It shows how interview materials can be used to improve our understanding of the cultural and interactional contexts in which health technologies are delivered.

Bloor M (1976). Bishop Berkeley and the adenotonsillectomy enigma: an exploration of variation in the social construction of medical disposals. *Sociology;* 10:43–51.

Bloor studied the way in which different ear, nose and throat (ENT) surgeons arrived at their decisions about how children who had been referred by their GPs for possible adenotonsillectomy should be treated. He carried out a series of observations in the outpatient clinics of 11 different ENT specialists. He found that all the specialists used routine assessment practices, which made their disposal decisions unproblematic. They were able to build up pictures of the clinical signs, symptoms and circumstances of each patient, which clearly indicated how the patient in question should be treated.

Bloor analysed his data using a modified form of analytic induction. He identified the decision rules and search procedures, which made up each specialist's routine assessment practices. Decision rules were idiosyncratic rules of thumb, which dictated the particular symptoms and signs that the individual specialist treated as the minimum criteria for making a disposal decision (e.g. deciding whether or not to list the child for surgery). Search procedures were the standard procedures used by individual doctors to obtain the information necessary to apply such decision rules.

Bloor found that the eleven specialists differed in both the decision rules and the search procedures that they used to arrive at their decision about a child's treatment. For example, specialists differed in the physical examinations carried out and the weight they gave to examination findings, compared to history taking, when arriving at their decisions. They also varied in the ways in which they took histories from the patient/parent and the way in which they made use of such histories. The routines that specialists used were also found to vary with the age of the child.

Bloor's study illustrates how the careful and systematic analysis of observational data can be used to identify the taken-for-granted, everyday practices of health professionals, which have significant implications for the use of resources within the health service. As Bloor commented, the differences that he observed between the routine practices of individual surgeons were unlikely to be visible to the surgeons themselves who had very limited opportunities to observe the practices of their fellow specialists.

Bloor M (1994). On the conceptualization of routine decision making: death certification as a habitual activity. In: Bloor M, Taraborelli P, editors. Qualitative studies in health and medicine. Aldershot: Avebury. p. 96–109.

In this study, Bloor was concerned to understand the factors which lead to the variation in death certification practices which had been observed in previous research. The study was carried out in two locations in Scotland. Using data from the General Register of Office for Scotland, Bloor identified 31 doctors who had completed death certificates for at least seven people within the previous 6 months. Twenty-eight of these doctors agreed to take part in the study.

Bloor used a combination of semi-structured interviews and a dummy death certification exercise in this study. In the interviews, the doctors were asked about their death certification practices. Areas covered in the interviews included doctors' opinions about the layout of death certificates and the kinds of diagnostic terms which were commonly used. Other topics included dealing with relatives and the tracing of deaths back to comparatively remote underlying causes. These interviews were audio tape-recorded and transcribed. Additional interviews

were carried out with non-clinical personnel, such as the senior deputy fiscal (who carries out duties similar to those of the coroner in England), the sergeant in charge of a big police enquiry department and senior nursing officers at homes for the elderly, who played an important role in the processing of death.

The dummy death certification exercise involved each doctor in filling out a series of six dummy death certificates in response to standardised case summaries. A control group (who had not been interviewed) were also asked to fill in the certificates based on the case summaries, for comparative purposes. The certificates were coded in the same way as genuine death certificates in order to identify the underlying cause of death identified by each doctor for each summary.

The dummy death certificates were analysed to reveal any differences between doctors in the underlying cause of death which the doctors examined. Bloor found substantial differences in the cause of death that different doctors identified on the basis of the same case summaries. Bloor was less interested in documenting this variation (which had been found in previous research) than in examining the way in which such variation came about. In particular, he noted that death certification was, on the whole, a solitary activity which individual doctors carry out in isolation from their professional peers. As such it was not generally regulated by reference to common guidelines. This did not mean, however, that it was a merely arbitrary or random activity. Each doctor's death certification practice could be seen to be routinised and underpinned by certain moral imperatives, such as sparing relatives unnecessary grief, which were commonly referred to by the doctors in the study.

This study illustrates the strength of qualitative methods in describing the process by which a recognised phenomenon (in this case the variation in death certification practices between doctors) occurs and in suggesting ways in which undesired practices may be changed through intervention. Understanding such processes is important in illuminating the mechanisms which obstruct the implementation of new approaches.

Buckholdt D, Gubrium J (1979). Caretakers: treating emotionally disturbed children. Beverley Hills, CA: Sage.

This is a participant observation study of an institution for treating emotionally disturbed young people. The researchers were interested in discovering how people see and understand their lives in care and treatment institutions. They focused particularly on 'professional practice', studying the practical and social processes through which various staff (including social workers, child care workers, special education teachers etc.) came to see the children in the institution as more or less disturbed.

The study involved prolonged and intensive participant observation in one institution (*Cedarview*). The two researchers participated in all aspects of life at the treatment centre for a period of one year. Observations were carried out in a range of settings within the institution. These included the classrooms, the residences, conferences about individual students and individual counselling sessions. The researchers' roles changed during the course of the study. For example, in their early observations of classrooms, the researchers were relatively passive observers but, as time passed, they began to participate more fully and eventually, in some cases, came to play the role of assistant teacher. Observational data were supplemented by documentary data (e.g. children's files) and impromptu interviews with staff.

In selecting *Cedarview* as the site of their observation, the authors bore in mind a number of considerations. They were anxious to study an institution in which treatment was actually carried out. They also wanted to study a setting that was recognised as being of high quality, so as to avoid the possibility that their findings would be attributed to the shortcomings of the staff or facilities of a particular centre. *Cedarview* met these selection criteria.

The authors used field notes to record their observations. They involvement of two observers allowed discussion of observation and interpretations. The researchers also showed their notes to staff members from time to time and were reassured to discover that they found nothing 'extraordinary' in what had been written. Buckholdt and Gubrium presented this as confirmation that the data they collected reflected the everyday life of the institution.

This study illustrates the way in which extended participant observation can uncover the social processes that are operating within institutions and which may, because of their taken-for-granted nature, be opaque to the staff and clients within that setting. While, as the authors acknowledged, the restriction of a study to a single setting does limit the generalisability of the findings, the study nevertheless identifies a number of significant

social processes which could be further studied in other settings.

Dingwall R, Murray T (1983). Categorization in accident departments: 'good' patients, 'bad' patients and 'children'. ***Sociology of Health and Illness*****;5:127–48.**

The starting point for this study was earlier research by Roger Jeffrey, in which he had studied the ways in which patients in A&E departments were categorised by staff and the consequences which this had for their treatment within the department (Jeffery, 1979). Jeffery reported that patients in the department were categorised as either 'good' or 'rubbish' and that those who were categorised as rubbish (in particular those whose complaints were trivial or self-inflicted) were treated less sympathetically than good patients.

As part of a wider study of decision-making by agencies responsible for the care and protection of children, Dingwall and Murray observed the A&E department in a teaching hospital in the South East of England. These observational data were supplemented by interviews with medical and nursing staff in this and three other English A&E departments. The departments studied served catchment areas that differed in terms of socio-economic factors and geographical position and were selected to provide a social and economic cross-section of the country outside the major conurbations.

Dingwall and Murray used their data to examine the extent to which Jeffery's analysis could be applied to children in A&E departments. They found that children did not fit into Jeffery's framework. Children routinely broke the 'rules' in ways in which, following Jeffery's analysis, would normally lead to categorisation as a rubbish patient. In particular, most of the injuries presented by children were the direct result of their own behaviour and did not restrict the child's activities. Children were often perceived by the medical staff as uncooperative. However, in spite of all this, children were not designated as rubbish and did not attract the kind of punitive treatment which Jeffery had reported in the case of adults. Nor were their parents treated punitively as they might have been if they had been considered as proxies for their children's shortcomings as good patients.

By studying the 'negative case' of children in A&E departments, Dingwall and Murray were able to modify Jeffery's analysis. They argued that Jeffery's analysis of patient categorisation in terms of 'good' and 'bad' patients is not sufficiently sophisticated to encompass the social processes involved. While patients may be identified as 'deviant' or 'conforming' at intake, they will not necessarily stay in these initial categories. Thus, while children may initially be allocated to the deviant category, they are likely to be re-categorised as 'special' since children are not normally considered responsible for their actions. They also observed that both good and bad patients can be re-categorised as 'interesting' if they offer staff an opportunity for learning or to practice experts skills.

On the basis of these observations, Dingwall and Murray proposed that the work of A&E departments is organised within one of three alternative frameworks: the bureaucratic, the special and the clinical. Where patients were conforming and their condition was of no particular interest, the bureaucratic framework operated. Where the patient was deemed deviant in some way, the special frameworks came into play and these differed depending upon whether the patients were seen as responsible for their problems. The clinical framework operated where the patient was deemed clinically interesting in some way.

This is an example of the way in which qualitative research can be used to test out hypotheses drawn from earlier studies in new settings. By identifying a 'negative' case, in terms of Jeffery's original formulation, Dingwall and Murray were able to revise the original analysis in a way which makes it more sophisticated and inclusive.

Gantley M, Davies D, Murcott A (1993). Sudden infant death syndrome: links with infant care practices. ***BMJ;*****306:16–20.**

The starting point for this study was the epidemiological evidence of national and ethnic variation in SIDS. In particular, the authors were concerned to identify the factors which might contribute to the low incidence of SIDS in the Asian population in the UK.

Gantley and co-workers carried out 60 interviews with mothers, living in Cardiff, whose babies were less than one year old. One-third of the sample were of Bangladeshi origin, one-third were Welsh working class and one-third were Welsh middle class. Care was taken to match the Welsh and Bangladeshi women for age of baby, type of accommodation and area of the city. The women in the Welsh and Bangladeshi groups were served by the same health professionals.

The aim of the qualitative interviews used in this study was to document the beliefs and child-rearing practices of women in different cultural groups, which might throw light on their different incidences of SIDS. Each mother was asked to describe a 'day in the life of' her baby. The interviewee was given the initiative in terms of the pace of the interview and the order in which topics were discussed. However, the interviewer had a checklist of topics to be covered and occasionally asked for clarification and/or expansion. The interviews, which lasted about one hour, were audio-tape recorded and fully transcribed.

The authors organised their analysis around five themes (living patterns, family networks, times and dates, sleeping patterns, and inter-dependence), which were identified from the data. A number of possible risk factors for SIDS, such as the amount of stimulation around sleeping babies and the likelihood that the baby will sleep in a separate room, were identified.

The study illustrates the hypothesis generating potential of qualitative interviews and the complementarity of qualitative and quantitative research. Qualitative interviews were used to identify possible mechanisms underlying the variation in SIDS incidence, which had been noted in large-scale epidemiological studies.

Heritage J, Sefi S (1992). Dilemmas of advice: aspects of the delivery and reception of advice in interactions between health visitors and first time mothers. In: Drew P, Heritage J, editors. Talk at work. Cambridge: Cambridge University Press.

This is a detailed study of the way in which advice is given by health visitors and received by first-time mothers during domiciliary visits. Advice-giving is the ostensible purpose of these visits and the success of such visits depends largely upon the uptake of such advice by parents. The authors used the techniques of CA to investigate the implications of different patterns of advice giving for uptake of such advice.

The data were drawn from eight initial visits made by five different health visitors to first-time mothers in their own homes. The visits were audio-tape recorded and fully transcribed using a set of transcription conventions that indicate the length of pauses, overlapping talk, particular emphasis and so on. From these eight interviews, the authors identified 70 advice giving sequences. These sequences were defined as those in which the health visitor described, recommended or otherwise forwarded a preferred course of future action.

These advice-giving sequences were analysed using techniques drawn from CA. They were categorised as those that were health visitor-initiated and those that were mother-initiated. Mother-initiated sequences were further sub-divided into those where a direct request for advice was made and those where advice was elicited by describing an untoward state of affairs. In both cases mothers tended to present themselves as knowledgeable and competent. Health visitor-initiated sequences ranged across a continuum from those in which the health visitor engaged in relatively elaborate attempts to establish a 'problem' in advance, to those in which advice was delivered to a completely unprepared recipient.

The authors went on to compare the reception of advice by mothers in mother- and health visitor-initiated advice-giving sequences. They assigned mothers' reception of advice to one of three categories: marked acknowledgement, unmarked acknowledgement and assertions of parental knowledge or competence. In one further case the advice was openly rejected.

Heritage and Sefi concluded that only one of the three categories of parental response (the marked acknowledgement) implied a fully-fledged acceptance of the advice given. The other two categories were interpreted as implying either active or passive resistance. Heritage and Sefi found that such resistance was particularly likely to occur where advice-giving was health visitor-initiated, with three-quarters of all health visitor initiated sequences receiving either unmarked acknowledgement or assertion of parental knowledge/competence.

This study is an example of the way in which detailed analysis of what actually happens in interactions between health professionals and patients or clients can be used to assess the effectiveness of everyday professional practice.

Morgan M, Watkins CJ (1988). Managing hypertension: beliefs and responses to medication among cultural groups. *Sociology of Health and Illness*;10(4):561–78.

This study was concerned to improve understanding of patient responses to hypertension and advice about its management. The authors noted evidence from earlier studies that compliance

with hypertension medication is low. Previous explanatory studies had sought to compare compliant and non-compliant patients in an attempt to identify the characteristics of each group. Morgan and Watkins adopted a different approach, seeking to elicit patients' own beliefs, concerns and patterns of behaviour and examining so-called 'compliant' behaviour in detail.

The authors recruited 60 hypertensive patients from 15 general practices in Lambeth. All the interviewees were aged between 35 and 55 years and were of manual occupational class. The sample was made up of equal numbers of men and women and half were 'white' and half were West Indian. All the interviewees had been diagnosed for at least one year and none had concomitant chronic disease.

The data were collected using qualitative interviews, lasting approximately one hour, in patients' own homes. These interviews consisted largely of open-ended questions and patients were encouraged to talk freely about their beliefs, concerns and practices. They were asked about having high blood pressure, their medication practices in general and, in more detail, their drug taking in the previous week.

The analysis presented in this paper focuses upon two topics: the perceived causes of hypertension and drug-taking behaviour. The most frequently cited cause, across both ethnic groups, was 'tension, worry or stress'. Other causes mentioned by respondents included heredity, diet and being overweight. The authors found differences between 'white' and West Indian respondents in terms of drug taking behaviour. Almost all 'white' respondents claimed to take their medication as prescribed, compared to less than half of the West Indians. A significant form of non-compliance reported by the West Indian respondents involved 'leaving off' their medication for a period of time. Various reasons were given for this practice, including concerns about drug dependency, avoiding mixing medication and alcohol, and reluctance to take medication when they did not feel unwell. The authors also noted that many of the West Indian respondents combined herbal remedies and prescribed medication without informing their GPs.

The phenomenon of non-compliance with hypertensive medication has been documented in quantitative studies. This study exploits the flexibility of qualitative methods to develop our understanding of the beliefs and practices that are associated with such non-compliance.

Silverman D (1981). The child as a social object: Down's syndrome children in a paediatric cardiology clinic. ***Sociology of Health and Illness*****;3:254–74.**

This study is concerned with the way in which disposal decisions are organised and announced to parents in a paediatric cardiology clinic. Silverman noted that the clinic policy varied depending upon whether or not the child patient had Down's syndrome in addition to a serious heart defect. In the case of a heart defect alone, the bias was towards intervention, whereas in the case of a heart defect in conjunction with Down's syndrome, the bias was towards non-intervention. However, contrary to what one might expect, there was rarely any conflict between parents and medical staff in relation to the decision not to operate on children with Down's syndrome. Indeed, in these cases, it was ostensibly the parents who usually made the decision that an operation was inappropriate in the case of their particular child. The focus of this paper is upon the mechanics of persuasion: the practices and procedures which were routinely used in the clinic to induce parents to agree with the clinic's policy in a relatively smooth and trouble-free fashion.

The analyses presented here were based on transcripts of 34 audio-tape recorded consultations, which took place over a one year period, with the same doctor, in one cardiology clinic. In 12 of these consultations, the child in question had Down's syndrome in addition to a major heart defect and, in the other 22, the child had a major heart defect only. In analysing these data, Silverman carried out a detailed comparison of the 'normal' and Down's syndrome consultations. He divided the consultations into ten possible stages and compared what occurred in each stage in the two groups.

Silverman reported a number of systematic differences between the Down's syndrome consultations and the others. Many of these had the effect of shifting the discussion of the Down's syndrome child's condition away from the clinical sphere, where intervention might be called for. Whereas the focus of the consultations with 'normal' children was upon their state of health, the doctor turned the discussion of Down's syndrome children to social functioning and the possible negative effects of intervention.

In consultations with the parents of children with Down's syndrome, the doctor referred the decision about surgical intervention to parents, whereas, in other consultations, the doctor took responsibility

for making the decision and then justified that decision to the parents. Silverman considered the functions which deferring to consumer choice served when the patients were children with Down's syndrome. He concluded that, among other things, it freed the doctor from appearing to play God, while at the same time making it very likely that the eventual decision would reflect the clinic's policy of non-intervention.

This study demonstrates the usefulness of detailed observation of everyday clinical practice. By paying close attention to what was actually happening within the clinic, Silverman was able to identify the processes through which a potentially controversial clinical policy was negotiated in a trouble-free fashion. His analysis raises significant questions about the nature of 'consumer choice' in medical settings. This study illustrates the importance of integrating a small number of 'negative cases' (those which do not fit the normal pattern) into the analysis, rather than simply treating them as statistically insignificant.

Silverman D (1984). Going private: ceremonial forms in a private oncology clinic. *Sociology*;18(2):191–204.

In this paper, Silverman compared private and NHS oncology clinics. The central question which he addressed was whether or not patients in a private oncology clinic bought a distinctive product. The starting point for the analysis was Strong's analysis of the 'ceremonial order' of the clinic. Strong found that interactional patterns in NHS clinics were based upon the same professional dominance and 'politeness ethic' that were characteristic of private settings. Silverman sought to test out this finding in oncology clinics.

The data presented in this paper were drawn from observations of three oncology clinics. These data were subjected to both quantitative and qualitative analyses. Two of these clinics were held in NHS teaching hospitals and the third was a private clinic held by one of the two oncologists observed in the NHS clinics. This allowed direct comparison of a private and an NHS clinic held by the same doctor and where only the methods of payment differed.

One hundred and forty-six consultations (104 NHS, 42 private) were observed, in 25 clinic sessions (16 NHS, nine private), with nine doctors. The age and gender of NHS and private patients were broadly similar, but they differed in terms of occupation (only one private patient was a manual worker) and country of origin (nine of the 42 private patients were foreign nationals).

Silverman found that the private and NHS consultations were broadly similar in terms of both professional dominance and the 'politeness ethic' identified by Strong. However, there were also substantial areas of difference between the two types of clinic. There were differences between the physical surroundings. The private clinic took place in lavish surroundings and Silverman concluded that these affected the nature of the interaction. Private consultations were longer than NHS consultations and, possibly as a result, it was more likely that non-medical matters would be discussed in private consultations. The service was more personalised in the private clinic and personal data were more likely to be treated as belonging to the particular patient. Private patients were found to be more likely to control the agenda of the consultation and to raise topics that were rarely observed at NHS consultations. Patients at the private clinic were observed to be more mobile in the consultation room. For example, they were more likely to move around the room when the doctor was absent.

In this study, Silverman demonstrated how qualitative and quantitative measures can be combined within the same study. He used quantitative methods (such as measuring the length of each consultation and counting the numbers of statements initiated by the patient) to identify differences between private and NHS consultations. He then used more detailed qualitative methods to carry out analyses of the behaviour and talk which contributed to these differences.

Silverman D, Bor R, Miller R, Goldman E (1992). 'Obviously the advice is then to keep to safer sex': advice-giving and advice reception in AIDS counselling. In: Aggleton P, Davies P, Hart G, editors. AIDS: rights, risks and reason. London: Falmer.

The focus of this paper is upon the organisation and reception of advice in HIV counselling sessions. It draws upon the earlier work of Heritage and Sefi (1992) in which they studied the delivery and reception of advice in interactions between health visitors and first-time mothers. As such, Silverman and co-workers are taking hypotheses developed in one setting and seeking to test them out in another.

The data were drawn from counselling sessions in ten different centres in England, Trinidad and the USA. More than 100 counselling sessions were audio tape-recorded and transcribed using CA conventions. Silverman and co-workers analysed the data using the framework proposed by Heritage and Sefi (1992). Four possible sequences for advice giving were identified:

- stepwise entry consisting of health professional enquiry, a client response indicating a problem, a request for specification by the professional, a specification by the client and, finally, advice giving
- as above but with no request for specification because client indicated how the problem was dealt with
- no client statement of how the problem was dealt with and no professional request for specification
- professional initiated advice without client giving a response indicating a problem.

In line with Heritage and Sefi's findings, Silverman and co-workers found that the majority (61 out of 100) of advice giving sequences consisted of a counsellor enquiry followed directly by advice giving, with no client input or problem specification. They also found a relationship between the way in which advice giving sequences were set up and the likelihood of uptake. Uptake was found to be less likely where advice was given without any attempt to elicit a perceived problem from the client and more likely when a stepwise approach was used.

Silverman and co-workers concluded that stepwise sequences are a more effective method of advice giving, increasing the likelihood of uptake. When this method is used, clients learn relevant information as well as the skill to determine what is appropriate for themselves and their partner.

This study is an example of the way in which the generalisability of the findings from one study can be demonstrated by using its findings to generate hypotheses to be tested in another setting or context. In this case qualitative methods are used to test such hypotheses. Detailed study of the interaction between professionals and clients allowed the researchers to examine the mechanisms by which the effectiveness of advice giving may be enhanced or compromised.

Stimson G, Webb B (1975). Going to see the doctor. London: Routledge.

In this research, Stimson and Webb analysed the general practice consultation, with particular reference to the way in which consultations are managed as a social process. The focus was upon studying the consultation from the perspective of the patient. Rather than restricting their study to the face-to-face encounter in the doctor's surgery, they treated the periods before and after such encounters as an important part of the consultation process.

Stimson and Webb used a range of data collection methods in this study. They interviewed patients, recorded conversations with practice staff, observed consultations between doctors and patients, and elicited descriptions of the family and social life of the study patients from one GP. Observations were also made in six pharmacies and 20 patients filling prescriptions at pharmacies were interviewed. Focus groups were held at which 20 women were encouraged to talk about illness, doctors and medicine, and essays about going to see the doctor were collected from school children. In addition, interviews with doctors who were taking part in another study and impromptu conversations which arose during the study were also used.

The data were collected in a South Wales town and most came from two general practices within that town. The practices were largely self-selected insofar as the GPs were sympathetic to the study.

Stimson and Webb argued that patients' behaviour prior to the consultation was best understood as a 'rehearsal' in which patients prepared themselves to manage the consultation as effectively as possible. During the consultation, both doctors and patients engaged in strategic negotiations in which both parties attempted to influence the other and self-presentation was crucial. After the consultation, patients reinterpreted what had occurred, attempting to fit this into their own frameworks. The outcome of this reinterpretation was seen as crucial for the patient's future behaviour and for compliance, in particular.

An important finding from this study was the discrepancy between the interview accounts that patients gave of their consultations and the actual observations, which the researchers made of such consultations. The interview accounts took the form of dramatic presentations, in which the patient was cast as hero and the doctor as incompetent. In the observations, the patients were found to be passive and reluctant to challenge or question doctors. Rather than treating such discrepancies as evidence of the untrustworthiness of interview data, Stimson and Webb focused upon what patients were actually **doing** when they tell

such 'atrocity stories'. They argued that the stories are best understood as a "vehicle for making the patient appear rational and sensible and for redressing the imbalance between patient and doctor" (97).

This study illustrates the usefulness of combining a range of data collection methods in one study. Such mixed methods are not used here as a way of testing the validity of one method by comparing its findings with that of another. Rather, they are used to complement one another and to avoid the trap of drawing superficial conclusions from a single source of data.

Strong PM (1979a). The ceremonial order of the clinic: parents, doctors and medical bureaucracies. London: Routledge & Kegan Paul.

This research concerns meetings between doctors and parents of children who were sick or undergoing medical inspection. The focus was upon the doctor–patient relationship and, in particular, upon the ceremonial aspects of meetings between parents and doctors. Strong examined the social form of the occasion and the kinds of identity which are tacitly claimed by each party and conferred upon the other. He was less concerned with issues such as the medical aspects of the encounter or the feelings, opinions or perspectives of those involved.

The primary data were collected through observation of outpatient clinics in two hospitals. One of these was in Scotland and the other in USA. Strong emphasised the importance of preserving the independence of data and analysis and providing readers with sufficient data to check the validity of the conclusions drawn.

Written case notes were taken of each observation. While acknowledging the shortcomings of this method, Strong argued that it had the advantage of allowing him to study a large number of cases (1000+). This was important insofar as it led to the inclusion of a number of deviant cases in his analysis. Such deviant cases allowed him to extend and modify his analysis. Strong considered the possibility that his presence in the clinics had distorted the data but concluded that this risk was minimal since the doctors were likely to be preoccupied with their everyday tasks and because it was not unusual for third parties to be present during encounters between doctors and patients.

Strong described his method of analysis as a cross between Glaser's "constant comparative method of qualitative analysis" (Glaser, 1964) and analytic induction. He generated propositions from the data and then these were constantly tested and re-formulated in the light of new data. He also used a 'split halves' technique. Hypotheses were generated and refined using half the data. These were subsequently tested in relation to the second half of the data.

Strong's main findings were that most NHS patients were dealt with in what he called a bureaucratic fashion, a rather bland and impersonal style of practice, which tried to avoid explicit moral judgements or prescriptions about the patients' or parents' own responsibility for their condition. This became important in his challenge to the then fashionable thesis of medicalisation, which claimed that doctors were increasingly intruding into areas of life that were more moral than technical. This claim was based on the writings of advocates of an expansion of medicine's role but Strong's direct observation showed that it had little foundation in routine everyday practice.

Voysey M (1975). A constant burden: the reconstitution of family life. London: Routledge & Kegan Paul.

Voysey's research is a vivid illustration of some of the problems that are raised when researchers attempt to use interview data (whether qualitative or quantitative) as representing reality in some kind of unproblematic way.

The research was concerned with the effect of having a disabled child upon family life. Voysey herself set out with the intention of using interviews with the parents of disabled children to give such parents "a chance to make themselves heard" (61). The initial choice of qualitative methods for this study was based upon a resistance to trying to measure phenomena "whose dynamics are not yet fully understood" (66).

Four 'semi-structured' interviews were carried out, over the course of approximately one year, with the parents of 13 children who had 'relatively serious' and 'probably permanent' physical or mental disabilities. Nine further families participated in some interviews but were unable or unwilling to complete the full series of interviews. The interviews, which were audio tape-recorded, lasted approximately one and a half hours. All the children were newly diagnosed and had both parents living at home. The interviews covered five topics: the onset of the child's disability,

encounters with medical agencies, the disability, family life, and encounters with others.

In carrying out the research, Voysey came to see that there were significant problems in using the data as the 'voice' of the parents of disabled children. She realised that the interview accounts she was collecting said more about cultural expectations about the phenomenon of having a disabled child than they did about the experience itself. This did not render the interview accounts invalid. Rather it had implications for the way in which such accounts could be used. Voysey opted to treat the interviews as accounts, which were produced by the parents in specific circumstances for specific purposes, and which reflected the societal values surrounding disabled children. Although she had begun with the intention of writing what we would now call a standpoint ethnography, she came to realise that this was an impossible task.

Appendix 2

Problems associated with the use of computerised record keeping

From Greatbatch, Heath and Luff (unpublished manuscript)
Greatbatch and co-workers described the problems associated with the use of the computerised records as opposed to the paper records are as follows:

> Paper documents are ecologically mobile. The traditional A5 medical records and folder, and of course the small A6 prescription pad used in general practice can be easily moved around the desk and the consulting room. This means, for example, that the prescription pad and pen can be placed between doctor and patient on the desk so that only a minor shift in orientation is required for the doctor to shift gaze from one to the other. With the computer, however, shifts in gaze and bodily orientation are more marked, especially if the screen is placed away from the patient at the centre or far end of the desk. Similar issues also apply in the case of the medical record cards, which can be held in the hand while standing, placed over the knee while sitting facing the patient, propped on the corner of the desk while conducting an examination, or held in front of the patient in order to facilitate a collaborative readings. The ecological mobility of the paper documents provides doctors with far greater flexibility than is currently available in the computing systems being used within primary health care.
>
> Paper documents co-locate reading and writing. You are able to read where the pen touches the paper. In the case of the computer system, however, the standard keyboard and monitor separate spatially and visually the domain in which text is controlled or entered from the domain in which it is read. Reading is spatially fragmented from both the manipulation of text and writing. when for example a practitioner wishes to scan a patient's medical history, the devices that are use to scroll through text are separated from the text itself, and even with relatively experienced users' we find they inevitably glance at the keyboard both before scrolling and sometimes during. Entering information namely typing is more complicated still, And again, even with those who are familiar with the use of the keyboard, we find that during course of documenting information they successively glance between the keyboard the text they are entering. The physical separation of the activity demanded by screen and keyboard, and the ongoing shifts in visual orientation it necessitates, appear to undermine even the more the sophisticated 'users' ability to delicately co-ordinate reading and writing with the ongoing and contingent demands of the interaction.
>
> The computer system 'responds' and 'displays' options. Whereas paper documents are inanimate and provide no response to action which are performed through the media, the computer is designed to set options and to indicate to the user when those actions are complete. Moreover, the as a consequence of a previous action by the doctor, we can find the image on the computer screen undergoing radical change some moments later. Doctors often need to closely monitor the operation of the computer to ensure that the appropriate responses have been elicited, as well as to enable them to co-ordinate their own actions with the movement of the cursor along the prompt line and other changes on the screen. Doctors are also be required to attend output messages, such as requests for clarification, corrections to inputs and warning 'beeps'. Doctors do not have to monitor paper documents to discern whether information has been 'accepted'.
>
> The computer sets prespecified patterns and sequences of input. Once practitioners has begun a particular activity and started along the prompt lines, they have little control over the order in which the information is entered. This, coupled with the blinking cursor, perhaps predisposes doctors to continue and complete a screen based activity, despite the potential demands which arise concurrently within the interaction with the patient. In contrast, paper records and prescription pages, whilst embodying certain constraints on the layout and position of textual information, place no constraint on pattern in which information is entered, or even necessarily the sequential structure of the actions which document that information. Whilst, for example, it might be, or at least was, unusual with paper records for the doctor to list prescription details prior to detailed the presenting problem and diagnosis, as long as a small gap at the beginning of the entry is left open, there is no reason at all why the information should not be entered in whatever sequence takes the whim of the doctor. More importantly perhaps, it is the responsive sequence of prompts in entering text which may well undermine the doctors ability to remain sensitive to the patient whilst simultaneously using the computer.
>
> The computer systems which are currently used within primary health care have reduced the doctors ability

to simultaneously participate in discussions with the patient whilst documenting and retrieving medical information. The 'constraints' of system use appear particularly pertinent discussions which involve topics which are tangential to the computational 'task' at hand. The use of the system, both when reading and writing, appears to demand a commitment, an involvement, in the computational 'task' at hand, which is unparalleled by the use of the traditional paper records and prescription pad.

Health Technology Assessment panel membership

This report was identified as a priority by the Methodology Panel.

Acute Sector Panel

Chair: Professor John Farndon, University of Bristol †

Professor Senga Bond, University of Newcastle-upon-Tyne †

Professor Ian Cameron, Southeast Thames Regional Health Authority

Ms Lynne Clemence, Mid-Kent Health Care Trust †

Professor Francis Creed, University of Manchester †

Professor Cam Donaldson, University of Aberdeen

Mr John Dunning, Papworth Hospital, Cambridge †

Professor Richard Ellis, St James's University Hospital, Leeds

Mr Leonard Fenwick, Freeman Group of Hospitals, Newcastle-upon-Tyne †

Professor David Field, Leicester Royal Infirmary †

Ms Grace Gibbs, West Middlesex University Hospital NHS Trust †

Dr Neville Goodman, Southmead Hospital Services Trust, Bristol †

Professor Mark P Haggard, MRC †

Mr Ian Hammond, Bedford & Shires Health & Care NHS Trust

Professor Adrian Harris, Churchill Hospital, Oxford

Professor Robert Hawkins, University of Bristol †

Dr Gwyneth Lewis, Department of Health †

Dr Chris McCall, General Practitioner, Dorset †

Professor Alan McGregor, St Thomas's Hospital, London

Mrs Wilma MacPherson, St Thomas's & Guy's Hospitals, London

Professor Jon Nicholl, University of Sheffield †

Professor John Norman, University of Southampton

Dr John Pounsford, Frenchay Hospital, Bristol †

Professor Gordon Stirrat, St Michael's Hospital, Bristol

Professor Michael Sheppard, Queen Elizabeth Hospital, Birmingham †

Dr William Tarnow-Mordi, University of Dundee

Professor Kenneth Taylor, Hammersmith Hospital, London

Diagnostics and Imaging Panel

Chair: Professor Mike Smith, University of Leeds †

Professor Michael Maisey, Guy's & St Thomas's Hospitals, London *

Professor Andrew Adam, UMDS, London †

Dr Pat Cooke, RDRD, Trent Regional Health Authority

Ms Julia Davison, St Bartholomew's Hospital, London †

Professor Adrian Dixon, University of Cambridge †

Mr Steve Ebdon-Jackson, Department of Health †

Professor MA Ferguson-Smith, University of Cambridge †

Dr Mansel Hacney, University of Manchester

Professor Sean Hilton, St George's Hospital Medical School, London

Mr John Hutton, MEDTAP International Inc., London

Professor Donald Jeffries, St Bartholomew's Hospital, London †

Dr Andrew Moore, Editor, *Bandolier* †

Professor Chris Price, London Hospital Medical School †

Dr Ian Reynolds, Nottingham Health Authority

Professor Colin Roberts, University of Wales College of Medicine

Miss Annette Sergeant, Chase Farm Hospital, Enfield

Professor John Stuart, University of Birmingham

Dr Ala Szczepura, University of Warwick †

Mr Stephen Thornton, Cambridge & Huntingdon Health Commission

Dr Gillian Vivian, Royal Cornwall Hospitals Trust †

Dr Jo Walsworth-Bell, South Staffordshire Health Authority †

Dr Greg Warner, General Practitioner, Hampshire †

Methodology Panel

Chair: Professor Martin Buxton, Brunel University †

Professor Anthony Culyer, University of York *

Dr Doug Altman, Institute of Health Sciences, Oxford †

Professor Michael Baum, Royal Marsden Hospital

Professor Nick Black, London School of Hygiene & Tropical Medicine †

Professor Ann Bowling, University College London Medical School †

Dr Rory Collins, University of Oxford

Professor George Davey-Smith, University of Bristol

Dr Vikki Entwistle, University of Aberdeen †

Professor Ray Fitzpatrick, University of Oxford †

Professor Stephen Frankel, University of Bristol

Dr Stephen Harrison, University of Leeds

Mr John Henderson, Department of Health †

Mr Philip Hewitson, Leeds FHSA

Professor Richard Lilford, Regional Director, R&D, West Midlands †

Mr Nick Mays, King's Fund, London †

Professor Ian Russell, University of York †

Professor David Sackett, Centre for Evidence Based Medicine, Oxford †

Dr Maurice Slevin, St Bartholomew's Hospital, London

Dr David Spiegelhalter, Institute of Public Health, Cambridge †

Professor Charles Warlow, Western General Hospital, Edinburgh †

* Previous Chair
† Current members

continued

continued

Pharmaceutical Panel

Chair: Professor Tom Walley, University of Liverpool †

Professor Michael Rawlins, University of Newcastle-upon-Tyne*

Dr Colin Bradley, University of Birmingham

Professor Alasdair Breckenridge, RDRD, Northwest Regional Health Authority

Ms Christine Clark, Hope Hospital, Salford †

Mrs Julie Dent, Ealing, Hammersmith & Hounslow Health Authority, London

Mr Barrie Dowdeswell, Royal Victoria Infirmary, Newcastle-upon-Tyne

Dr Tim Elliott, Department of Health †

Dr Desmond Fitzgerald, Mere, Bucklow Hill, Cheshire

Dr Felicity Gabbay, Transcrip Ltd †

Dr Alistair Gray, Health Economics Research Unit, University of Oxford †

Professor Keith Gull, University of Manchester

Dr Keith Jones, Medicines Control Agency

Professor Trevor Jones, ABPI, London †

Ms Sally Knight, Lister Hospital, Stevenage †

Dr Andrew Mortimore, Southampton & SW Hants Health Authority †

Mr Nigel Offen, Essex Rivers Healthcare, Colchester †

Dr John Posnett, University of York

Mrs Marianne Rigge, The College of Health, London †

Mr Simon Robbins, Camden & Islington Health Authority, London †

Dr Frances Rotblat, Medicines Control Agency †

Mrs Katrina Simister, Liverpool Health Authority †

Dr Ross Taylor, University of Aberdeen †

Dr Tim van Zwanenberg, Northern Regional Health Authority

Dr Kent Woods, RDRD, Trent RO, Sheffield †

Population Screening Panel

Chair: Professor Sir John Grimley Evans, Radcliffe Infirmary, Oxford †

Dr Sheila Adam, Department of Health*

Ms Stella Burnside, Altnagelvin Hospitals Trust, Londonderry †

Dr Carol Dezateux, Institute of Child Health, London †

Dr Anne Dixon Brown, NHS Executive, Anglia & Oxford †

Professor Dian Donnai, St Mary's Hospital, Manchester †

Dr Tom Fahey, University of Bristol †

Mrs Gillian Fletcher, National Childbirth Trust †

Professor George Freeman, Charing Cross & Westminster Medical School, London

Dr Mike Gill, Brent & Harrow Health Authority †

Dr JA Muir Gray, RDRD, Anglia & Oxford RO †

Dr Anne Ludbrook, University of Aberdeen †

Professor Alexander Markham, St James's University Hospital, Leeds †

Professor Theresa Marteau, UMDS, London

Dr Ann McPherson, General Practitioner, Oxford †

Professor Catherine Peckham, Institute of Child Health, London

Dr Connie Smith, Parkside NHS Trust, London

Dr Sarah Stewart-Brown, University of Oxford †

Ms Polly Toynbee, Journalist †

Professor Nick Wald, University of London †

Professor Ciaran Woodman, Centre for Cancer Epidemiology, Manchester

Primary and Community Care Panel

Chair: Dr John Tripp, Royal Devon & Exeter Healthcare NHS Trust †

Professor Angela Coulter, King's Fund, London *

Professor Martin Roland, University of Manchester *

Dr Simon Allison, University of Nottingham

Mr Kevin Barton, East London & City Health Authority †

Professor John Bond, University of Newcastle-upon-Tyne †

Ms Judith Brodie, Age Concern, London †

Dr Nicky Cullum, University of York †

Professor Shah Ebrahim, Royal Free Hospital, London

Mr Andrew Farmer, Institute of Health Sciences, Oxford †

Ms Cathy Gritzner, The King's Fund †

Professor Andrew Haines, RDRD, North Thames Regional Health Authority

Dr Nicholas Hicks, Oxfordshire Health Authority †

Professor Richard Hobbs, University of Birmingham †

Professor Allen Hutchinson, University of Sheffield †

Mr Edward Jones, Rochdale FHSA

Professor Roger Jones, UMDS, London

Mr Lionel Joyce, Chief Executive, Newcastle City Health NHS Trust

Professor Martin Knapp, London School of Economics & Political Science

Dr Phillip Leech, Department of Health †

Professor Karen Luker, University of Liverpool

Professor David Mant, NHS Executive South & West †

Dr Fiona Moss, Thames Postgraduate Medical and Dental Education †

Professor Dianne Newham, King's College London

Professor Gillian Parker, University of Leicester †

Dr Robert Peveler, University of Southampton †

Dr Mary Renfrew, University of Oxford

Ms Hilary Scott, Tower Hamlets Healthcare NHS Trust, London †

* Previous Chair
† Current members

Health Technology Assessment 1998; Vol. 2: No. 16

Copies of this report can be obtained from:

The National Coordinating Centre for Health Technology Assessment,
Mailpoint 728, Boldrewood,
University of Southampton,
Southampton, SO16 7PX, UK.
Fax: +44 (0) 1703 595 639 Email: hta@soton.ac.uk
http://www.soton.ac.uk/~hta

ISSN 1366-5278