

## Outcomes of social care for adults: developing a preference-weighted measure

A Netten, P Burge, J Malley, D Potoglou,  
A-M Towers, J Brazier, T Flynn, J Forder  
and B Wall



March 2012  
10.3310/hta16160

Health Technology Assessment  
NIHR HTA programme  
[www.hta.ac.uk](http://www.hta.ac.uk)





### **How to obtain copies of this and other HTA programme reports**

An electronic version of this title, in Adobe Acrobat format, is available for downloading free of charge for personal use from the HTA website ([www.hta.ac.uk](http://www.hta.ac.uk)). A fully searchable DVD is also available (see below).

Printed copies of HTA journal series issues 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 issue and for the rest of the world £3 per issue.

How to order:

- fax (with **credit card details**)
- post (with **credit card details** or **cheque**)
- phone during office hours (**credit card** only).

Additionally the HTA website allows you to either print out your order or download a blank order form.

### **Contact details are as follows:**

Synergie UK (HTA Department)  
Digital House, The Loddon Centre  
Wade Road  
Basingstoke  
Hants RG24 8QW

Email: [orders@hta.ac.uk](mailto:orders@hta.ac.uk)

Tel: 0845 812 4000 – ask for ‘HTA Payment Services’  
(out-of-hours answer-phone service)

Fax: 0845 812 4001 – put ‘HTA Order’ on the fax header

### **Payment methods**

#### *Paying by cheque*

If you pay by cheque, the cheque must be in **pounds sterling**, made payable to *University of Southampton* and drawn on a bank with a UK address.

#### *Paying by credit card*

You can order using your credit card by phone, fax or post.

### **Subscriptions**

NHS libraries can subscribe free of charge. Public libraries can subscribe at a reduced cost of £100 for each volume (normally comprising 40–50 titles). The commercial subscription rate is £400 per volume (addresses within the UK) and £600 per volume (addresses outside the UK). Please see our website for details. Subscriptions can be purchased only for the current or forthcoming volume.

### **How do I get a copy of HTA on DVD?**

Please use the form on the HTA website ([www.hta.ac.uk/htacd/index.shtml](http://www.hta.ac.uk/htacd/index.shtml)). *HTA on DVD* is currently free of charge worldwide.

---

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

# Outcomes of social care for adults: developing a preference-weighted measure

A Netten,<sup>1\*</sup> P Burge,<sup>3</sup> J Malley,<sup>1,2</sup> D Potoglou,<sup>3</sup>  
A-M Towers,<sup>1</sup> J Brazier,<sup>4</sup> T Flynn,<sup>5,6</sup> J Forder<sup>1,2</sup>  
and B Wall<sup>7</sup>

<sup>1</sup>University of Kent at Canterbury, Canterbury, UK

<sup>2</sup>London School of Economics and Political Science, London, UK

<sup>3</sup>RAND Europe, Cambridge, UK

<sup>4</sup>University of Sheffield, Sheffield, UK

<sup>5</sup>University of Bristol, Bristol, UK

<sup>6</sup>University of Technology, Sydney, Australia

<sup>7</sup>Accent, London, UK

\*Corresponding author

**Declared competing interests of authors:** none

Published March 2012

DOI: 10.3310/hta16160

---

This report should be referenced as follows:

Netten A, Burge P, Malley J, Potoglou D, Towers A-M, Brazier J, *et al.* Outcomes of social care for adults: developing a preference-weighted measure. *Health Technol Assess* 2012;**16**(16).

*Health Technology Assessment* is indexed and abstracted in *Index Medicus/MEDLINE*, *Excerpta Medica/EMBASE*, *Science Citation Index Expanded (SciSearch®)* and *Current Contents®/Clinical Medicine*.

The Health Technology Assessment (HTA) programme, part of the National Institute for Health Research (NIHR), was set up in 1993. It produces high-quality research information on the effectiveness, costs and broader impact of health technologies for those who use, manage and provide care in the NHS. 'Health technologies' are broadly defined as all interventions used to promote health, prevent and treat disease, and improve rehabilitation and long-term care.

The research findings from the HTA programme directly influence decision-making bodies such as the National Institute for Health and Clinical Excellence (NICE) and the National Screening Committee (NSC). HTA findings also help to improve the quality of clinical practice in the NHS indirectly in that they form a key component of the 'National Knowledge Service'.

The HTA programme is needed in that it fills gaps in the evidence needed by the NHS. There are three routes to the start of projects.

First is the commissioned route. Suggestions for research are actively sought from people working in the NHS, from the public and consumer groups and from professional bodies such as royal colleges and NHS trusts. These suggestions are carefully prioritised by panels of independent experts (including NHS service users). The HTA programme then commissions the research by competitive tender.

Second, the HTA programme provides grants for clinical trials for researchers who identify research questions. These are assessed for importance to patients and the NHS, and scientific rigour.

Third, through its Technology Assessment Report (TAR) call-off contract, the HTA programme commissions bespoke reports, principally for NICE, but also for other policy-makers. TARs bring together evidence on the value of specific technologies.

Some HTA research projects, including TARs, may take only months, others need several years. They can cost from as little as £40,000 to over £1 million, and may involve synthesising existing evidence, undertaking a trial, or other research collecting new data to answer a research problem.

The final reports from HTA projects are peer reviewed by a number of independent expert referees before publication in the widely read journal series *Health Technology Assessment*.

#### Criteria for inclusion in the HTA journal series

Reports are published in the HTA journal series if (1) they have resulted from work for the HTA programme, and (2) they are of a sufficiently high scientific quality as assessed by the referees and editors.

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 research reported in this issue of the journal was commissioned by the National Coordinating Centre for Research Methodology (NCCRM), and was formally transferred to the HTA programme in April 2007 under the newly established NIHR Methodology Panel. The HTA programme project number is 06/96/01. The contractual start date was in August 2008. The draft report began editorial review in March 2011 and was accepted for publication in August 2011. The commissioning brief was devised by the NCCRM who specified the research question and study design. The authors have been wholly responsible for all data collection, analysis and interpretation, and for writing up their work. The HTA editors and publisher have tried to ensure the accuracy of the authors' report and would like to thank the referees for their constructive comments on the draft document. However, they do not accept liability for damages or losses arising from material published in this report.

The views expressed in this publication are those of the authors and not necessarily those of the HTA programme or the Department of Health.

Editor-in-Chief: Professor Tom Walley CBE

Series Editors: Dr Martin Ashton-Key, Professor Aileen Clarke, Dr Tom Marshall, Professor John Powell, Dr Rob Riemsmas and Professor Ken Stein

Associate Editor: Dr Peter Davidson

Editorial Contact: [edit@southampton.ac.uk](mailto:edit@southampton.ac.uk)

ISSN 1366-5278 (Print)

ISSN 2046-4924 (Online)

ISSN 2046-4932 (DVD)

© Queen's Printer and Controller of HMSO 2012. This work was produced by Netten *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health.

This journal is a member of and subscribes to the principles of the Committee on Publication Ethics (COPE) (<http://www.publicationethics.org/>).

This journal may be freely reproduced for the purposes of private research and study and may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NETSCC, Health Technology Assessment, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

Published by Prepress Projects Ltd, Perth, Scotland ([www.prepress-projects.co.uk](http://www.prepress-projects.co.uk)), on behalf of NETSCC, HTA.

Printed on acid-free paper in the UK by the Charlesworth Group.

# Abstract

## Outcomes of social care for adults: developing a preference-weighted measure

A Netten,<sup>1\*</sup> P Burge,<sup>3</sup> J Malley,<sup>1,2</sup> D Potoglou,<sup>3</sup> A-M Towers,<sup>1</sup> J Brazier,<sup>4</sup> T Flynn,<sup>5,6</sup> J Forder<sup>1,2</sup> and B Wall<sup>7</sup>

<sup>1</sup>University of Kent at Canterbury, Canterbury, UK

<sup>2</sup>London School of Economics and Political Science, London, UK

<sup>3</sup>RAND Europe, Cambridge, UK

<sup>4</sup>University of Sheffield, Sheffield, UK

<sup>5</sup>University of Bristol, Bristol, UK

<sup>6</sup>University of Technology, Sydney, Australia

<sup>7</sup>Accent, London, UK

\*Corresponding author

**Objective:** The aim of this study was to develop a measure of social care outcome, an equivalent to the quality-adjusted life year (QALY) in health, which could be used in a range of circumstances.

**Design:** The project drew on previous and parallel work developing the Adult Social Care Outcome Toolkit and the national Adult Social Care Survey. We developed and tested an instrument designed to reflect service users' social care-related quality of life (SCRQoL) and tested it with 30 service users from a variety of user groups and 300 older home care service users. In parallel, we explored discrete choice experiment (DCE) and best-worst scaling (BWS) approaches to preference elicitation with 300 members of the general population, and cognitively tested these with service users. We also cognitively tested a computer-aided time trade-off (TTO) exercise using SCRQoL attributes with members of the general population. In the second phase, using the finalised instruments, BWS interviews were conducted with 500 members of the general population, TTO interviews with a follow-up sample of 126 of these respondents, and BWS interviews with 458 people using equipment services.

**Main outcome measures:** The final measure had eight domains: *personal cleanliness and comfort, accommodation cleanliness and comfort, food and drink, safety, social participation and involvement, occupation, control over daily life and dignity*. In addition to measuring current SCRQoL, the instrument includes questions used to establish service users' views of their 'expected' SCRQoL in the absence of services. The difference between a person's current and 'expected' SCRQoL provides an indicator of service impact.

**Results:** There was good evidence for the validity of the descriptive system and the validity of the current, expected and SCRQoL gain scales. The DCE and BWS approaches yielded similar results and, once introductions made clear, were understood by service users. BWS was used for the main stages, as it had technical and cognitive advantages. The computer-aided approach to TTO worked well, and respondents found questions acceptable and understandable. There were no substantive differences in the preferences of service users and the general population. The key domain was *control over daily life*, with the lowest and

highest levels strongly estimated in all models. After allowing for observable heterogeneity, service users' preferences appeared to be more closely associated with their own SCRQoL than with those of the general population. The consistency of the results with the results of a previous study allowed the final model to be based on the preferences of 1000 members of the general population. A formula based on the relationship between TTO and BWS values was estimated for a social care QALY, with '0' equivalent to 'being dead' and '1' being the 'ideal' SCRQoL state. Members of the population experienced significantly higher SCRQoL than service users.

**Conclusions:** Although further work is needed, particularly to develop an equivalent measure for informal carers and to explore the links with health QALYs, the measure has considerable potential. A number of methodological advances were achieved, including the first application of TTO in a social care context and use of BWS to establish service user preferences.

**Funding:** The National Institute for Health Research Health Technology Assessment programme.

# Contents

<b>List of abbreviations</b>	<b>vii</b>
<b>Executive summary</b>	<b>ix</b>
<b>1. Introduction</b>	<b>1</b>
Background	1
Aims and objectives	1
Research design	2
Conceptual basis of the measure	3
Conclusion	8
<b>2. Development of items</b>	<b>9</b>
Introduction	9
Previous development and review of Adult Social Care Outcome Toolkit	9
Analysis of previous versions of Adult Social Care Outcome Toolkit	10
Development of the new Adult Social Care Outcome Toolkit questions	12
Testing the revised Adult Social Care Outcome Toolkit instrument	12
Conclusions	17
<b>3. Testing validity</b>	<b>19</b>
Introduction	19
Methods	20
Description of sample	21
Current social care-related quality of life	22
Social care-related quality-of-life gain	29
Conclusion	30
<b>4. Preference elicitation development of methods</b>	<b>33</b>
Introduction	33
Eliciting preferences	33
Comparison of best–worst scaling and discrete choice experiment valuations	35
Service user cognitive testing	44
Time trade-off	46
Implications for the main stage design	48
<b>5. Development of utility weights</b>	<b>51</b>
Introduction	51
Design of the best–worst scaling experiment	51
Best–worst scaling results	58
Time trade-off study	79
Conclusion	82
<b>6. The measure</b>	<b>83</b>
Introduction	83
Population and service user item distributions	83
Utility weighting	85
Population and service user distributions of the measure	88

Validity and reliability	88
Limitations	90
Implications and potential future uses	92
Further work	92
Conclusions	94
<b>Acknowledgements</b>	<b>97</b>
<b>References</b>	<b>99</b>
<b>Appendix 1</b> Development of domains of social care-related quality of life	<b>107</b>
<b>Appendix 2</b> Analysis of previous versions of the Adult Social Care Outcome Toolkit	<b>109</b>
<b>Appendix 3</b> Meaning of response options by domain	<b>119</b>
<b>Appendix 4</b> Validity of revised <i>accommodation</i> and <i>food and drink</i> items	<b>123</b>
<b>Appendix 5</b> Final version of the Adult Social Care Outcome Toolkit instrument	<b>125</b>
<b>Appendix 6</b> Main best–worst scaling method preference study and follow-up samples	<b>135</b>
<b>Appendix 7</b> Domain-level weights of a collapsed three-option version of Adult Social Care Outcome Toolkit	<b>149</b>
<b>Appendix 8</b> Choice elicitation and model estimation procedure in best–worst scaling data	<b>155</b>
<b>Health Technology Assessment programme</b>	<b>161</b>



## List of abbreviations

ADL	activities of daily living
ASCOF	Adult Social Care Outcomes Framework
ASCOT	Adult Social Care Outcome Toolkit
ASCS	Adult Social Care Survey
BME	black and minority ethnic
BWS	best–worst scaling
CAPI	computer-aided personal interview
CASP	a scale for measuring quality of life in older people
CI	confidence interval
DCE	discrete choice experiment
df	degrees of freedom
EFA	exploratory factor analysis
EQ-5D	European Quality of Life-5 Dimensions
G-MNL	generalised multinomial logit model
GHQ	General Health Questionnaire
HRQoL	health-related quality of life
IADL	instrumental activities of daily living
IBSEN	the study evaluating the use of individual budgets
ICC	intraclass correlation coefficient
ICECAP	the ICEpop CAPability instrument from the Investigating Choice Experiments for the Preferences of Older People (ICEPOP) project. The instrument has been renamed ICECAP-O to distinguish it from the new ICECAP-A version for adults of any age
KMO	Kaiser–Meyer–Olkin
LA	local authority
LSE	London School of Economics
max	maximum
min	minimum
MNL	multinomial logit model
MOPSU	Measuring Outcomes for Public Service Users
OMEP	orthogonal main effects plan
OPUS	Older People's Utility Scale
OSCA	Outcomes of Social Care for Adults
PSS	Personal Social Services
PSSRU	Personal Social Services Research Unit
PWLD	people with learning difficulties
PWMH	people with mental health problems
PWPSI	people with physical and sensory impairments
QALY	quality-adjusted life year
QoL	quality of life
S-MNL	scale heterogeneity multinomial logit model
SC-QALY	social care quality-adjusted life year
SCRQoL	social care-related quality of life
SD	standard deviation
SE	standard error
SG	standard gamble
SPOW	Social Production of Welfare

TTO	time trade-off
UES	User Experience Survey
VA	visual analogue
VAS	visual analogue scaling

---

All abbreviations that have been used in this report are listed here unless the abbreviation is well known (e.g. NHS), or it has been used only once, or it is a non-standard abbreviation used only in figures/tables/appendices, in which case the abbreviation is defined in the figure legend or in the notes at the end of the table.

# Executive summary

## Background

There is an increasing policy focus on outcomes in the field of health and social care in England. In times of financial stringency, it is particularly important to be able to identify the impact of interventions in order to assess cost-effective alternatives and make the best use of limited resources. In the field of health, the quality-adjusted life year (QALY) is widely accepted as the measure of health outcome used to evaluate the cost-effectiveness of new health interventions. There has been no equivalent for social care, despite its growing importance as the population ages. Lack of such a measure has made it difficult to assess accurately the full impact of interventions intended to address both health and social care, to plan future resource needs for long-term care, and to move to outcome-based commissioning that makes best use of resources.

## Objectives

The study was designed to develop an equivalent measure to the QALY in health that would reflect the particular characteristics of social care and could be used in a range of circumstances to reflect the impact and value of social care interventions. The objectives were to:

- develop a measure of social care outcome that captures all relevant domains, has credibility in the social care community, has been cognitively tested and demonstrates good psychometric properties
- test a variety of approaches to establishing preference weights for methodological soundness and practical application with the general population
- investigate how applicable these methods are to a service user population
- conduct a population preference study
- develop a set of weights for calculating social care QALYs (SC-QALYs) and weighting measures of social care output
- explore how the views of service users differ from those of the general population, and establish alternative preference weights, if appropriate
- identify the factors that affect preferences among the general population and service users.

## Methods

The project comprised two phases. The first development and feasibility phase covered the first three objectives, and was used to inform the design for the second phase, in which preference weights for the final measure were estimated.

The development of the measure during the first phase included conceptual development, a focused literature review, analysis of data sets that incorporated previous versions of the measure, cognitive testing of items with 30 service users, and a survey of 300 older people using home care. We built on a number of previous studies on outcome measurement in social care, and the project both contributed to, and drew on, parallel work undertaken as part of the Measuring Outcomes for Public Service Users (MOPSU) project, which itself fed into the development of the Adult Social Care Outcome Toolkit (ASCOT). In addition, some refinements were made to the wording of two of the items in the instrument as a result of concerns about the distribution

of these items. These were cognitively tested with 25 service users as part of a study feeding into the development of the national Adult Social Care Survey (ASCS), and the final set of items was included in a pilot postal survey for the ASCS of 1364 people from all service user groups. An important element of the development phase of the work was the involvement of key stakeholders, including service users, local councils and experts in the field, to assist in ensuring the practicality and acceptability of the proposed measure.

The preference elicitation feasibility work involved testing alternative discrete choice approaches through a survey of 300 members of the general population, and cognitively testing these approaches in the interviews with 30 service users about the instrument. In order to explore the feasibility of anchoring the measure to the state of 'being dead', we cognitively tested the use of time trade-off (TTO) techniques with 19 members of the general population. A preference elicitation survey of 1000 members of the general population, undertaken as part of the MOPUSU project, was used to test the proposed best–worst scaling (BWS) design, with half of the sample using the provisional social care-related quality-of-life (SCRQoL) domains and levels developed for this study.

The second phase drew on this iterative work and, for the final version of the instrument, involved a main preference study of 500 members of the general population using BWS and a follow-up TTO exercise with a subsample of 126 people. We also conducted a preference study with 458 people who used equipment services to establish service user preferences, and undertook a follow-up study with a subsample of 101 of these people to test further the final version of the instrument.

## Results

### *Development of the Adult Social Care Outcome Toolkit measure*

Health outcome measures identify specific aspects of health-related quality of life in order to ensure that they are sensitive to the impact of health-care interventions. In addressing social care we used an equivalent – SCRQoL. Our review and analysis of previous measures of SCRQoL resulted in eight domains that were deemed to be relevant to a greater or lesser degree to all service user groups: *personal cleanliness and comfort, accommodation cleanliness and comfort, food and drink, safety, social participation and involvement, occupation, control over daily life* and *dignity*. All of these domains had been included in earlier versions of the measure, with the exception of *dignity*, intended to reflect the psychological impact of support and care on a service user's personal sense of significance. A further domain, which identified whether or not the individual was 'living in own home', was tested but rejected for the final measure.

Previous versions had defined three levels of 'need' in each domain ('no', 'low' and 'high'). As a result of our analyses and consultations, we decided that a fourth level should be added to make the measure more sensitive to differences within the 'no needs' option. Here, we aimed to reflect Sen's distinction between 'capabilities' and 'functionings'. The latter concept is based on the principle that society is not prepared for citizens to maintain such a poor level of functioning in any domain that there are health implications if their needs are not met. Once needs are met, it is also essential to identify capabilities: whether or not people are able to achieve their desired situation. Each domain has four levels, aimed to reflect as closely as possible:

- *Ideal* The preferred situation, where needs are met to the desired level.
- *No needs* Where needs are met, but not to the desired level.
- *Some needs* Where there are needs, but these do not have an immediate or longer-term health implication.

- *High needs* Where there are needs that have an immediate or longer-term health implication.

The long-term, compensatory nature of social care makes it particularly challenging to measure outcomes. But the fact that people are usually well aware of – and have often experienced – what their situation would be like in the absence of services does aid the study of social care outcomes, and enabled us to pursue a pragmatic approach to establishing outcomes in this area. Drawing on previous work, we incorporated questions for each domain that identified whether or not services had an impact on this aspect of their lives and, if so, their ‘expected’ situation in the absence of the service. These questions allow the estimation of three SCRQoL measures: *current* or experienced; *expected* in the absence of support and services; and *gain*, which provides an indicator of impact.

Cognitive testing refined the wording and confirmed that the responses to the ‘expected’ items appeared to function as intended across a range of service user groups. The testing of the *dignity* item suggested the inclusion of an additional item (not included in scoring the measure), to allow for the fact that some people have problems coming to terms with needing help at all. The interviews with a larger sample of older home care users showed that response rates were good for both the current and expected items, suggesting that the pragmatic approach to outcome measurement was feasible with this population.

There was good evidence for the validity of the descriptive system for ASCOT: each attribute seemed to capture a different aspect of SCRQoL and had the anticipated relationship with other variables capturing similar and dissimilar concepts. There was also evidence for the validity of the SCRQoL scales reporting the current situation and the service impact. However, the distribution of service users across the different levels of *accommodation* and *food and drink* suggested that we were not reflecting the range of situations very effectively. Subsequent revisions resulted in improved distributions in the ASCS pilot survey and a revised version was taken through to the second phase of the project.

### Development of preference elicitation methods

Establishing preference weights for our measure was particularly challenging because of the number of attributes involved. We investigated and compared the results of using a discrete choice experiment design and a BWS approach as alternative preference elicitation techniques, both statistically in a sample of the general population and cognitively with service users. The two techniques yielded similar results, and both were seen to be understandable and acceptable to service users. We decided to use the BWS approach for the second phase of the study, as it had a number of advantages:

- respondents are presented with all domains at the same time
- coefficients are estimated on a common scale
- it poses a smaller cognitive burden on respondents.

After some modification, a computer-aided approach to TTO was found to work well. Respondents found questions acceptable and understandable, and lack of respondent fatigue allowed us to increase the number of scenarios presented in the final main stage design.

### Population and service user preferences

A key question for the project was whether there were significant differences between the SCRQoL preferences of the general population and those of service users. The models showed no substantive differences in the preferences of the two samples. The models were consistent, both showing differences in values in the expected directions and, with a few minor exceptions, statistically significant differences between the coefficients for the levels within each domain. The

key domain was *control over daily life*, strongly estimated in both models, with the lowest and highest levels demonstrating the lowest and highest values for any domain. The BWS analysis of general population preferences was able to capitalise on the MOPSU project data collection conducted 1 year earlier, as there was remarkable consistency between models based on that study and models based on the data collected as part of this project. Recent methodological developments have demonstrated the importance of allowing for sample-level variance heterogeneity, which can be seen as variations in the levels of certainty with which different groups express their preferences. Models were used to take account of observable heterogeneity between the two samples and other subgroups, both in generating utility weights and identifying factors associated with preferences. After allowing for these, there were some differences in factors associated with preferences, with service users' preferences more closely associated with aspects of their own SCRQoL.

The mean values for the 64 sample states in the TTO exercise showed a linear association with the BWS estimated values for those states. Some states were estimated as worse than 'being dead', with a TTO score of  $<0$ .

### **The Adult Social Care Outcome Toolkit measure**

The final BWS weights were estimated based on the combined MOPSU and Outcomes of Social Care for Adults sample of 1000 members of the general population. A formula based on the TTO analysis is provided for converting these to weights that could be used for a SC-QALY. This yields a range of possible scores between  $-0.171$  and  $1$ , with '0' being equivalent to 'being dead' and '1' being the 'ideal' SCRQoL state.

In order both to familiarise respondents with the domains and to provide context for interpreting the results, all of those who took part in the BWS preference surveys were asked about their current SCRQoL state using ASCOT. This provided us with a useful comparison between the general population and service users. As we might expect, members of the general population reported significantly higher SCRQoL than service users [SC-QALY scores of  $0.86$  and  $0.73$ , respectively ( $p < 0.001$ )]. Moreover, the differences were reflected in the areas expected, with members of the general population more likely to experience the 'ideal' state and service users the 'no needs' or 'mustn't grumble' state. As we would hope, no respondent was currently experiencing a state that would be rated as bad as or worse than 'being dead'.

## **Conclusions**

Resources are always limited, but this is even more the case in the current financial climate. This, together with the policy emphasis on outcomes, makes it all the more important that we have good measures to reflect the value of social care interventions for those who use them. The ASCOT measure now provides the basis for a social care equivalent to the QALY, which can be used in a range of circumstances to reflect the impact and value of social care interventions. The policy emphasis on giving service users greater control is supported by the results, and we have a basis on which to compare the SCRQoL of service users and the general population, potentially providing evidence for an 'outcomes' basis on which to make difficult decisions about resources.

The study was the first to use TTO in a social care context. Anchoring the score to the state of 'being dead' provides the first step in generating evidence on the relative cost-effectiveness of health and social care interventions. In addition, our pragmatic approach to reflecting the impact of services has the potential to provide a basis for measuring the impact or 'value added' of social care.

There has been widespread interest in the measure since June 2010, when ASCOT was first made widely available through [www.pssru.ac.uk/ASCOT](http://www.pssru.ac.uk/ASCOT). The measure is included in annual national ASCSs, and has been proposed as an important element of the Transparency in Social Outcomes Framework, which should enhance its value by providing benchmark data about current SCRQoL states for a range of service user groups and contexts.

In terms of future research, we would recommend:

- extending the approach to include informal carers
- exploring the potential to link ASCOT to the European Quality of Life-5 Dimensions in estimating and comparing QALYs
- developing validated approaches to establishing ASCOT states for service users who have cognitive impairment and communication difficulties and who are living in their own homes
- further work to test the validity, reliability and sensitivity of the current expected and gain measures in a variety of contexts.

## Funding

Funding for this study was provided by the Health Technology Assessment programme of the National Institute for Health Research.





# Chapter 1

## Introduction

### Background

There is increasing policy focus on outcomes in the field of health and social care in England.<sup>1,2</sup> This reflects a wider emphasis on the importance of reflecting outcomes of publicly funded services, which has been evident for some time, for example in reflecting productivity of public services in National Accounts.<sup>3</sup> In times of financial stringency it is particularly important to estimate the impact of different interventions accurately if we are to identify cost-effective alternatives and make the best use of limited resources. In the field of health, the quality-adjusted life year (QALY) is widely accepted as the measure of health outcome used to evaluate the cost-effectiveness of new health interventions. There is no such equivalent for social care, which has fundamentally different objectives to health care, making it difficult to assess accurately the impact of interventions that are closely aligned with, and can have direct impacts on, the demand for health care. The Wanless review<sup>4</sup> used an early social care outcome measure to comment on future resource needs, but identified the need for much more development of outcome tools in this area. Ideally, commissioners and providers should have outcome information to draw on as a management tool for allocating resources to ensure good productivity and efficient use of those resources.

This report describes the results of a study designed to develop an equivalent measure to the health QALY, which reflects the particular characteristics of social care and can be used in a range of circumstances to reflect the impact and value of social care interventions. This first chapter describes the aims and objectives of the study, provides a broad overview of the design and describes the conceptual basis of the measure. In *Chapter 2* we describe the basis for, and development of, the items that make up the instrument, and in *Chapter 3* we outline how we tested the validity of the measure. *Chapter 4* discusses how we tested a variety of approaches to preference elicitation and developed the methods used for the basis of the utility weights. The results of this work are reported in *Chapter 5*. The report ends with an overview of the final measure, discusses the potential uses and value of the measure, and suggests future work needed.

### Aims and objectives

The aim of the study was to generate a preference-weighted measure of social care outcomes that could be used in a variety of circumstances, including social care evaluations, cost-utility analyses, weighted output of government-funded social care for National Accounts,<sup>3</sup> and policy analyses.

The objectives were to:

- develop a measure of social care outcome that captures all relevant domains, has credibility in the social care community, has been cognitively tested and demonstrates good psychometric properties
- test a variety of approaches to establishing preferences for methodological soundness and practical application with the general population

- investigate the applicability of these methods to a service user population
- conduct a population preference study
- develop a set of weights for calculating social care quality-adjusted life years (SC-QALYs) and weighting measures of social care output
- explore how service users' views differ from those of the general population, and establish alternative preference weights if appropriate
- identify the factors that affect preferences among the general population and service users.

## Research design

The project comprised two phases. The first covered development and feasibility work to inform the design for a main phase in which preference weights for the final measure were derived from a study of both the general population and service users. The project undertook and drew on a number of different data collections, summarised in *Table 1* and described below.

There were two strands to the development and feasibility work, conducted in parallel:

- the development and psychometric testing of the measure itself
- exploration of the feasibility of different approaches to establishing preferences.

The aim was to refine and develop the measure, while testing a number of elicitation approaches with the general population. The results of these exercises could then feed into testing the measure and piloting preference elicitation with service users. This allowed us to evaluate both the measure itself and the preference elicitation methods.

The development of the measure included conceptual development, a focused literature review, analysis of data sets that incorporated previous versions of the measure, cognitive testing of items with service users, and a survey of older people using home care services (see *Table 1*, and *Chapters 2* and *3*). The preference elicitation feasibility work involved testing alternative discrete choice approaches through a survey of the general population and cognitively testing these approaches with service users. In addition, we wanted to explore the feasibility of anchoring the measure to the state of 'being dead', permitting equivalent cost–utility analyses as those used in health care. To this end, we cognitively tested the use of time trade-off (TTO) techniques with respect to social care-related quality-of-life (SCRQoL) attributes with members of the general population (see *Table 1* and *Chapter 4*).

We built on previous work in the field of outcome measurement in social care, and the project both contributed to, and drew on, other work that fed into the development of the Adult Social Care Outcome Toolkit (ASCOT) ([www.pssru.ac.uk/ASCOT](http://www.pssru.ac.uk/ASCOT)). The measure drew on earlier attempts to develop a measure of social care outcome for older people,<sup>5</sup> which was then extended through a number of studies to be applicable to a wider group of social care service users. The extension to other groups was undertaken partly as a result of an interest in developing an approach to developing outcome-based quality weightings for publicly-funded outputs for the purposes of National Accounts,<sup>3</sup> in which the Personal Social Services Research Unit (PSSRU) took the lead on developing an approach for Personal Social Services (PSS).<sup>6</sup> This work led to a Treasury-funded study led by the Office for National Statistics (ONS),<sup>7</sup> as part of which the ASCOT was developed, drawing on a study of care homes,<sup>8</sup> low-level services<sup>9</sup> and a preference study.<sup>10</sup> The preference study drew on the development work undertaken as part of this study and provided a helpful test of the measure during its developmental stage (see *Chapters 3* and *4*). In addition, development work for the national Adult Social Care Survey (ASCS), which now includes the ASCOT measure, provided other opportunities for testing the measure (see *Chapter 3*).<sup>11</sup>

The second phase drew on this iterative work and, for the final version of the instrument, involved a main preference study of the general population using the best–worst scaling (BWS) approach, and a follow-up TTO study with a subsample of this population. We also conducted a preference study of people who use equipment services, and a follow-up study with a subsample of these to test further the final version of the instrument (see *Table 1* and *Chapter 5*).

An important element of the work was the involvement of local councils, and other authorities in the field were recruited as ‘critical friends’ to the project. In the early stages of the work, to assist in ensuring the acceptability of the approach being developed, they were invited to comment on our thinking with respect to the conceptual basis of the measure and were later kept informed on its practical application.<sup>12</sup> The councils also provided access to service users who were interested in participating in various stages of the study through the annual User Experience Surveys (UESs) that councils are required to conduct.<sup>13</sup> Service users were also consulted through the user and carer group at the PSSRU, London School of Economics (LSE).

Ethical permission was obtained from the Social Care Research Ethics Committee (SCREC) for all stages of the work that involved service users. The Association of Directors of Adult Social Services (ADASS) supported the study, and research governance approval was obtained from all participating local councils. For elements of the study that involved only members of the general population, ethical approval was obtained from the University of Kent ethics committee. All participants were given gift vouchers to thank them for their involvement.

## Conceptual basis of the measure

Before we can develop a measure of outcome, we need to be clear about exactly what it is we are intending to measure. We have discussed the conceptual basis of outcome measurement in social care in more detail elsewhere.<sup>14–16</sup> Here, we summarise what we mean by social care, describe the

**TABLE 1** Data collections undertaken and drawn on during the study

Project	Nature of the sample	Sample size	Purpose
OSCA	People who use services (all service user groups) <i>See Chapter 2</i>	30	Cognitive testing of ASCOT measure items, cognitive testing of DCE and BWS questions
OSCA	Older home care service users ( <i>see Chapter 3</i> )	301	Psychometric testing of the ASCOT measure
ASCS development study <sup>8</sup>	People who use services (all service user groups)	25	Cognitively testing new versions of problematic ASCOT items
ASCS pilot study <sup>9</sup>	People who use services (all service user groups)	1364	Checking distribution of final version of all ASCOT items
OSCA	General population ( <i>see Chapter 4</i> )	300	Pilot preference elicitation study (comparing BWS and DCE)
OSCA	General population ( <i>see Chapter 4</i> )	19	Cognitively testing TTO technique with ASCOT measure
MOPSU preference study <sup>10</sup>	General population ( <i>see Appendix 6</i> )	513	BWS preference elicitation survey comparing three- and four-level versions of ASCOT measures
OSCA	General population ( <i>see Appendix 6</i> )	500	BWS preference elicitation survey of final ASCOT measure
OSCA	General population ( <i>see Appendix 6</i> )	126	TTO preference elicitation survey of final ASCOT measure
OSCA	Users of equipment services ( <i>see Appendix 6</i> )	458	BWS preference elicitation survey of final ASCOT measure
OSCA	Users of equipment services ( <i>see Appendix 6</i> )	100	Follow-up interviews using final ASCOT instrument

DCE, discrete choice experiment; MOPSU, Measuring Outcomes for Public Service Users; OSCA, Outcomes of Social Care for Adults.

implications of that for what we wanted to be able to reflect in our measure, and briefly consider the implications of the objective of a measure that can be seen as a SC-QALY equivalent.

### Social care

Social care is a term used in the UK to describe a range of long-term care activities, including providing help with personal hygiene, dressing and feeding, as well as help with shopping, getting out and about, socialising, and keeping the home tidy and clean. Social care is usually provided in response to needs arising from physical or sensory impairments, learning difficulties and mental health problems, including those associated with older age.<sup>17</sup> It is provided in people's homes, in day centres and care homes, and includes support purchased by individuals themselves through personal budgets or direct payments.

Most people using social care services have conditions that involve a permanent (and often declining) loss of functional ability, for example neurological conditions, arthritis and most intellectual disabilities. In these situations, the primary aim of social care interventions is to compensate a person for their lost functional ability, rather than try to restore it. Therefore, interventions focus primarily on maintaining service users' quality of life (QoL) by helping them with – or enabling them to accomplish – the types of tasks and activities associated with daily living described above. Frequently, because of the nature of a person's condition, this type of help is required continuously, often on a daily basis, until a person dies. For many service users, this means that the service is often an integral part of their life; indeed, and particularly in cases where the service helps with very personal tasks, the way the care is delivered (its process) can have a significant effect on the user's well-being.<sup>15,18</sup>

These characteristics of social care have important consequences for how we think about outcomes. Qureshi and Nicholas<sup>19</sup> identified three types of outcomes relevant to social care interventions: 'change' outcomes, which are rare and tend to be associated with rehabilitation; 'maintenance' outcomes, which are more common; and 'process' outcomes, which are related to how services are delivered and are important in their own right because of the ongoing and personal nature of much social care. To these, we might add 'prevention', which is increasingly being seen as important.<sup>20,21</sup> For each outcome type, it is important to ask 'of what?' For the most part, the answer is about well-being or QoL: maintaining, improving or preventing deterioration in well-being.

Social care is not unique in having the objective of improving QoL. Health, housing and other public services also have the goal of improving or maintaining QoL. If we use too broad a measure of QoL, we may fail to pick up on effective social care interventions because of the confounding effects of these and other such factors. The *extra-welfarist* approach addresses the problem by identifying the outcomes agreed by decision-makers as important.<sup>22-25</sup> For example, the government could decide that the aim of the health service is solely to improve population health. So, although health services have effects that go beyond people's health, the outcome of health care would be assessed only in these terms.

Extra-welfarist health outcome measures identify specific aspects of health-related quality of life (HRQoL) in order to ensure that they are sensitive to the impact of health-care interventions. Such measures have proved very powerful in generating an evidence base in health. Although social care is related to health care, its compensatory nature is fundamentally different, so HRQoL measures will likely be insensitive to the impact of social care interventions. Moreover, if we accept an extra-welfarist approach that puts weight on improving population health then we might conceive of an equivalent goal for social care. We might lack a familiar goal such as health improvement in social care, but government policy has indicated that the goal for social care is improving *well-being*.<sup>26</sup> In order to reflect the impact of social care with respect to well-being, we need to define and measure an equivalent to HRQoL: SCRQoL.

### **Social care-related quality of life**

If we are to define an equivalent to HRQoL for social care, we need to be more specific about what it is that social care does and aims to do. The Social Production of Welfare (SPOW) framework<sup>27,28</sup> puts individuals and their care networks at the heart of the production process. This draws on household economics<sup>29,30</sup> and the capabilities framework, put forward by Sen<sup>31,32</sup> as an alternative to standard welfare economics. Household economics propose that we are all in the business of producing welfare for ourselves and others in our households by using our resources, goods and services to produce what Sen defines as 'functionings' – such as social contact or being well fed – from which we directly derive utility or well-being. Impairment of a person in the household results in changes in the household production processes and can result in drawing others, including people, usually family members who are not resident in that household, into a 'care network'. Social care services and support become necessary when the resources of individuals and their care networks are insufficient to produce necessary levels of 'functioning' for network members' well-being, whether that of the person with impairments or of those caring for him/her.

From an extra-welfarist perspective, what are regarded as legitimate functionings for interventions or 'SCRQoL' can be deduced by observation of what services do, and consultation with those in receipt of services and support. The nature of the SPOW means that the impact of social care services will be experienced both by people with impairments and those caring for them. Informal carers can be represented as both providers of care and beneficiaries of social care services.<sup>33</sup> Provision specifically targeted on carers tends to take the form of providing respite from the caring task, substituting for the role that they are playing in the care network. The biggest impact, however, will come from ongoing services that have the well-being of the individual with impairment as their primary objective. These services will generate outcomes for both individuals and their carers, whose utility and well-being are inter-related.<sup>34</sup>

While acknowledging this, our focus here is on the impact on the primary recipient of social care services, whatever the source or nature of the care. In general, publicly funded services for individuals with impairment in England are primarily involved in the production of nutrition, personal care and safety, and, to a lesser extent, employment, social contact and participation, and household cleanliness and comfort. However, the focus of services is culturally defined, and varies both over time and across client groups.<sup>4,27</sup> For example, in England, home help services for older people used to be focused primarily on household cleanliness and comfort, with some personal care. In the 1990s, there was a policy shift away from what was seen as 'low-level' household support to more intensive services focused on personal care and safety.<sup>5</sup> Recently, there has been a move to more emphasis on 'prevention', social inclusion and on giving people choice and control.<sup>26,35,36</sup>

This shift in emphasis to choice and control has implications beyond the description of SCRQoL. Sen<sup>31</sup> argues that utility (or experienced QoL derived from functionings) is not the sole object of value; rather, it is capability – understood as the substantive opportunities an individual has to be, or to do, a range of things – that is the prime object of value. The importance of the capability-functioning distinction is that it differentiates, for example the person who does not go to the cinema out of choice from the person who does not go because he/she cannot for health or social reasons. Sen<sup>31</sup> argues that the latter person experiences a loss of utility associated with their lack of choice. This distinction is not recognised where functionings are the focus of analysis, but where capabilities are the focus it is recognised in a reduced capability set for the person who is restricted in what he/she can do. Although Sen<sup>31</sup> developed this theory in the context of poverty, it is apposite in the social care context, as people with social care needs frequently face barriers that constrain their choices.<sup>37,38</sup>

Sen's argument<sup>31</sup> for capabilities as the focus of measurement is supported by the social model of disability and the shift in policy in this area.<sup>38</sup> Traditionally, social care has been concerned with meeting needs, which could be represented as particular functioning states (such as being well-fed or safe). The social model of disability places contextual barriers in the foreground, giving them precedence over individual impairments, to emphasise that disability is a consequence of the environment rather than an intrinsic characteristic of a disabled person.<sup>39</sup> In keeping with this view, policies for younger disabled people have tended to focus on 'enablement' by, for example, changing the built environment and giving disabled people greater flexibility and freedom in their daily lives. In England, the emphasis has increasingly been on broadening opportunities for people with disabilities and developing 'independence', 'choice' and 'control'.<sup>26,35,36</sup> This more aspirational agenda has been reflected in policy documents, such as *Putting people first: a shared vision and commitment to the transformation of adult social care*<sup>20</sup> and, more recently, *A vision for adult social care: capable communities and active citizens*.<sup>40</sup> This focus on choice and control encourages us to aim to measure what people *can* do, rather than what they *actually* do, across all aspects of SCRQoL. Thus, when thinking about basic aspects of QoL, such as personal grooming, outcomes are not just about whether or not people have their personal care needs met but whether or not they can dress as they would like. Similarly, when considering occupation, we should be able to identify not solely whether or not people are employed or doing activities at all, but whether or not they are involved in work or activities that they value or enjoy.

### Measuring social care outcomes

Clearly, key to ensuring that any measure is a valid and sensitive measure of social care outcomes is establishing that all relevant domains of SCRQoL are covered, and that the way these domains are described reflects the objectives of social care in terms of capabilities and functionings. We describe our approach to establishing these domains of the measure in *Chapter 2*. A further issue is how the measure might be used in practice to establish outcome: the contribution of any given intervention to SCRQoL.

In many fields the standard approach is to measure the QoL state of a person before and again after an intervention is used, with the pre-test score acting as a proxy for what would have happened in the absence of the intervention. However, this approach assumes both that there is a clear start and end point to the intervention, and that the type of outcome observed displays (it is hoped) positive change. But social care interventions tend to be ongoing, so there is often no clear start or end point. In addition, people rarely approach services in a high-need state with no support, and if they did so it would be not only impractical, but also unethical, to ask them to participate in research before addressing those needs. This means that we cannot directly observe the SCRQoL state in the absence of the intervention, and the 'before' measurement tends to be taken when a person is already in receipt of social care. It is also likely that, over time, there will be an exacerbation of the underlying impairment and therefore functional ability of a person. As functional ability declines, the resources required to maintain their SCRQoL are greater, and it can become much harder to compensate them fully for their loss of functioning. Good long-term social care should maintain people's QoL, flexibly responding in the face of deteriorating or varying levels of impairment. With a 'before and after' measurement of SCRQoL, we would therefore expect to see either no change in SCRQoL where QoL has been maintained, or a decline in SCRQoL where it has not been possible to fully compensate the person for his or her declining functional ability. These factors mean that 'before' measures act as a poor proxy for 'absence of intervention'.

As with any evaluation, research designs (such as randomised controlled trials), observational data and analytical techniques can be used to disentangle and compare marginal effects, whether one intervention is superior to another for a specified group of people with particular

characteristics. Such designs that enable comparisons across interventions are clearly critical in this case to help interpret the observed lack of improvement in 'before and after' measurements. Such designs are very expensive, however, and are rarely conducted in social care. This means that there is a lack of data with which to inform decisions about targeting resources in a cost-effective manner.

The ongoing nature of most social care provides us with an advantage in terms of evaluating the QoL in the absence of services: people themselves daily face the possibility that the help and support on which they rely may not be forthcoming. For example, a care worker may be late or even not turn up at all. Because of these possibilities, social care service users are arguably uniquely well-qualified to identify for themselves what their SCRQoL would be in the absence of services.

In previous work,<sup>5</sup> and as part of ASCOT, we have therefore been developing a pragmatic approach to establishing outcomes, which builds on the nature of social care to establish an estimate of the full benefit of interventions. In the interview version of ASCOT, this involved asking people to estimate, for each domain, what their situation would be in the absence of the intervention and then comparing this estimate to their current state to generate a measure of SCRQoL gain from the intervention. Elsewhere,<sup>16</sup> we describe the rationale for the approach in more detail. In *Chapter 2* we describe the cognitive testing of the approach, and in *Chapter 3* we examine the evidence for validity.

The ongoing nature of social care interventions and the underlying conditions that services and support help to overcome mean that adaptation behaviour may be a particular issue when we are trying to evaluate social care. Generally speaking, where people have adjusted their expectations downwards when adapting to their circumstances (negative adaptation), we would expect people with low capability to rate their situation as better than if they had not adapted.<sup>41,42</sup> Similarly, people who have adapted to having a higher level of capability (positive adaptation) might find this situation less fulfilling than if they had not adapted. Adaptation is likely to mean we find a smaller effect of services than if adaptation did not occur: in the absence of services, service users who have negatively adapted might be expected to rate their QoL as higher than without adaptation. And with services that improve people's capability, service users might rate their QoL as lower than they would have without positive adaptation. The difference with adaptation is lower than the difference without adaptation.

In measuring the utility associated with (different levels of) capability, most QoL measures will be susceptible to this adaptation effect. Theoretically, we might argue that this problem is avoided if we are able to measure capability directly, but this would have to be done in an objective way so that respondents, in rating their capability, had no opportunity to allow their own preferences and expectations to colour their response. Designing the descriptive content of such an instrument would seem virtually impossible in practice. Potentially, a set of highly specific and verifiable metrics could be used (e.g. regarding social contact, we might measure the number, duration and types of contacts people made), but the set would need to be extremely large to comprehensively reflect all aspects of capability that might be affected by services. For more abstract aspects of capability, such as having control over daily life, it is hard to imagine how a set of objective metrics might be specified.

Given these challenges and the lack of a robust method to measure the extent of adaptation, we acknowledge that inevitably our measure will reflect (in so far as adaptation occurs) a person's adapted utility. We return to this issue in *Chapter 6*.

### Quality-adjusted life year equivalence

We have stated that one of our objectives was to develop a measure that could be used as a QALY equivalent. This requires that our measure is preference weighted, reflecting the relative importance of different aspects of SCRQoL and thus the value of the different states described by the measure. Preference weighting raises the issue of *whose* preferences, those of the general public or those of service users? There is an ongoing debate on this issue.<sup>43</sup> Seeking the general public's preferences has the important advantage that the preferences of members of the general public will be largely free of the influence of or adaptation to the circumstances of service users, particularly where those circumstances are affected by services. In other words, experience of a long-term condition and of how services support people with care needs is likely to shape the preferences of service users.

The disadvantage of using general population rather than service user preferences is that non-service users have to hypothesise what a situation of reduced functioning and capability is like. As Dolan and Kahneman<sup>43</sup> argue, general population respondents might focus on the short-term implication of transitioning into a poor functioning state, the transitional loss in utility.

We can also argue these points from a normative perspective. In particular, it might be regarded as important to establish preferences for the general population on the grounds that the population pays for health care through taxes, and decisions about funding (ultimately) are made at the ballot box.<sup>44</sup> However, there is also an argument that whose preferences should be addressed depends on why those preferences are being sought<sup>45</sup> and, in the field of social care, service users' views increasingly play a central role in policy and practice development. It is therefore important to establish if their views are systematically different to those of the general population and, if so, in what way.

In addition to preference weighting, we argue that if we can establish a common anchor and scale, this would allow economic evaluation of the impact of health and social care interventions, facilitating a broader perspective when evaluating the most effective use of health and social care resources. This would mean that interventions whose effects can be fully captured by a SCRQoL measure can be evaluated in a similar manner to those currently evaluated using a HRQoL one. Thus, the cost per social care-related QALY gained could be estimated and compared with the threshold(s) used by decision-makers. The base anchor used for QALYs is the state of 'being dead'. This is potentially challenging in the context of social care where mortality historically has rarely been identified as a relevant outcome but is clearly worth investigating to provide a more widely applicable and useful measure.

### Conclusion

The project was designed to establish a preference-weighted and ideally anchored generic measure of social care outcomes that could be used for a variety of purposes. The first task was to clarify exactly what it is we are attempting to measure. Drawing on the concepts identified above, in the next chapters we now turn to the development and validation of the instrument.



## Chapter 2

# Development of items

### Introduction

As noted in *Chapter 1*, we did not start from scratch in developing the measure: we were building on previous work and indicators that had been developed specifically to measure outcomes in social care. We made use of previous studies to review the domains and levels, both empirically and in the light of our conceptual thinking, which had moved on since the earlier measures had been developed. Although service users had been involved in previous work, early versions were based on the perspective of those observing (e.g. researchers), commissioning and providing social care. We consulted with service users through general advice (the LSE PSSRU user and carer advisory group), a review of the literature<sup>16</sup> and in the cognitive testing stages of this work and linked studies.<sup>46</sup>

In this chapter, we describe the reviews of the previous measures and our conclusions in terms of the domains, levels and wording requirements needed for the development of the individual items for our instrument. We then describe the results of the cognitive testing of those new items taken through to the next validation stage.

### Previous development and review of Adult Social Care Outcome Toolkit

The ASCOT measure was originally based on the Older People's Utility Scale (OPUS), a preference-weighted measure of social care outcomes for older people.<sup>5</sup> A Delphi exercise, asking respondents to focus on the key domains of outcome for older people, identified five domains (*food and drink, personal care, safety, social participation and involvement, and control over daily living*). Drawing on the literature and expert opinion, the OPUS measure was later extended to adults aged < 65 years, and included the domains *occupation, accommodation cleanliness and comfort, caring for others and living in own home*.<sup>47</sup> These reflected both the concerns of social care for younger age groups, such as parenting, and the objective of moving beyond 'key' domains to a more comprehensive measure covering all aspects of social care. A single question was used to reflect each domain, with the exception of *living at home*, which was simply recorded, reflecting whether the person was living in their own home or in a care home setting. The wording of the individual items was cognitively tested and changed as part of development work for a survey of younger adults.<sup>46</sup> A pilot preference study was conducted, providing provisional preference weights anchored to a financial attribute using discrete choice experiments (DCEs).<sup>10</sup> This version of the measure has been used in a number of studies,<sup>4,46,48</sup> including a major evaluation of the policy of individual budgets, a new form of consumer-directed support, in England.<sup>49</sup>

As part of the Outcomes of Social Care for Adults (OSCA) review of the conceptual basis for the measure, we noted that the ongoing and intimate nature of most social care interventions meant that care itself became a part of people's lives, which meant that the process of care-giving is important for people's QoL. The literature review found 15 studies in which aspects of process were identified as important to people's QoL, including such factors as being valued and treated

with dignity and respect,<sup>49–54</sup> being listened to/having a say,<sup>49–51</sup> having kind staff with whom one can have good relationships<sup>55–58</sup> and receiving a service that is both responsive and reliable.<sup>59–61</sup> For the most part, we would expect these aspects of process to be reflected in the SCRQoL domains. For example, people report that the reliability of a service affects their sense of control<sup>62</sup> and may have consequences for having needs met in other domains. We would expect the impact of abuse and neglect to be picked up primarily in people's sense of personal safety. However, some effects of the process of delivering care were unlikely to be captured, in particular those that have an impact on the psychological well-being of service users or their sense of personal dignity. We therefore decided to include in our measure an additional *dignity* domain, which captured the effect of the process of care-giving on a service user's sense of self-esteem. *Appendix 1* shows a table summarising the development of the domains through the different studies.

Conceptual work undertaken as part of the MOPUS study<sup>14</sup> (Measuring Outcomes for Public Service Users) had identified the importance of distinguishing between reflecting Sen's 'capabilities' and 'functionings'.<sup>31,32</sup> Wording of early versions of the measure had tended to reflect 'functioning' states most closely. One version of ASCOT developed for low-level interventions had focused on reflecting capabilities.<sup>9</sup> We wanted our measure to reflect the full range from the very fundamental level where functioning levels are so low they could lead to mental and physical health implications, through to 'capability' states, in which people have real choice and their circumstances can be seen to reflect the objectives of the 'personalisation' agenda in social care.<sup>11,40</sup>

## Analysis of previous versions of Adult Social Care Outcome Toolkit

At the time of the review for this study, the instrument had most recently been used in two studies: (1) a postal UES in 2007 of 2228 people with physical and sensory impairments (PWPSI) aged 18–64 years receiving help to enable them to live in their own homes and (2) in an evaluation of the use of individual budgets in social care, in which interviews were conducted with 959 people from all service user groups (the study evaluating the use of individual budgets – IBSEN). The data sets from these studies were analysed to explore empirically, first, whether there were any problems with the items as phrased and, second, the choice of domains. The data sets and analysis are described in detail in *Appendix 2*. Here we summarise the results and implications for the development of the measure.

In both data sets the domains included:

- control over daily life
- personal cleanliness and comfort
- food and nutrition
- accommodation
- safety
- social participation
- occupation.

In addition, the data sets included an item on 'caring for others', but we did not include this in our analyses as it was not seen as relevant by a very high a proportion of service users. In the IBSEN study, each item had three levels, reflecting:

- *No needs* Where any needs experienced are minor or the type of variation in experience people might normally expect in the absence of impairment.

- *Low or some needs* Where there are needs that lead to lower levels of well-being, but that are not likely to have any long-term consequences if unmet.
- *High needs* Where the needs are of such a level of severity or number that mental or physical health problems are likely to ensue in the short or longer term.

In the UES self-completion version, an additional level was included that was intended to distinguish whether 'no needs' was due to those needs already being met by services.

The response rates for the items in both data sets were generally good. However, the number of missing data was higher for the *occupation* and *social participation* items, suggesting that these could be improved. Response rates also varied across client groups; they were particularly good for PWPSI, and poor for people with learning difficulties (PWLD) and people with mental health problems (PWHM). It is possible that this pattern arose because the phrasing of the items used in both studies was developed in cognitive interviews with PWPSI.<sup>46</sup>

Across both of the data sets, the majority or the largest proportion of respondents chose the 'no needs' response option. [It should be noted that the distribution of some of the items (*control over daily life, safety and food and nutrition*) was better in the IBSEN than in the UES sample, which could be explained by the lack of the second 'service help' level in the IBSEN data set. This level could have been 'taking' respondents from the lower response options as well as the 'no needs' response option.] In addition, for the *accommodation cleanliness and comfort, food and nutrition* and *personal care* items, very few people in either data set chose the lowest response option. On both counts, this situation is desirable, as it implies that services are successful in compensating people for their impairments. However, it also means that the distribution of scores for any scale composed of these items is likely to be highly skewed, with most people to be found towards the high SCRQoL end of the scale and very few people towards the low SCRQoL end. The concentration of respondents at the high end of the measure could make the measure insensitive among populations with high SCRQoL. As we assumed that we would be unlikely to observe many people in poor SCRQoL situations, given the existence of services and the availability of help from other sources (family, friends and privately paid-for help), it is important to have high sensitivity at the high SCRQoL end of the scale.

In both data sets a similar relationship was observed between pairs of items. The *social participation* and *occupation* items had strong correlations, as did the *personal cleanliness and comfort* item with both *food and nutrition* and *accommodation cleanliness and comfort*. The strong correlation may indicate some conceptual overlap in these domains, which might cause problems in the preference elicitation study. In the case of the *personal cleanliness and comfort, food and nutrition* and *accommodation cleanliness and comfort* items, this explanation seemed unlikely, as the wording of the domains and the ideas the domains represent do seem conceptually distinct. A more likely explanation of the strong correlations noted is that they are all areas of life for which poor outcomes would have severe health implications, and service users (and service providers) are likely to express a preference towards having needs in these areas met. In other words, the strong correlation is a result of the similarity in users' preferences and the response of services to needs in these areas. There is more similarity in the concepts expressed by the *social participation* and *occupation* domains, and it was possible that there was some conceptual overlap between them. However, these two areas tend not to be priorities for social care services, so it may have been this perspective that was causing the strong correlation between these two items. Factor analysis for both data sets seemed to indicate that the items formed a weak unidimensional scale, with many of the items having high unique variances, suggesting that the items were conceptually distinct.

## Development of the new Adult Social Care Outcome Toolkit questions

The findings from the theoretical and literature review of the previous ASCOT instrument identified the need to develop a new *dignity* domain, and to ensure that Sen's<sup>31</sup> concept of capability was incorporated in the measure. From the empirical analysis we concluded there was a need to develop greater sensitivity at the high SCRQoL end of the scale by attempting to split the 'no needs' response option into two options: one representing the ideal situation, the other in which there are no real needs but the person considers the situation to be just adequate rather than completely satisfactory. In addition, the empirical work identified the need to improve the wording of the *social participation* and *occupation* domains, as well as to improve all the domains so that they were more applicable to older people, PWLD and PWMH. In summary, therefore, the literature and empirical review of the previous version of the measure suggested developing a version of ASCOT composed of nine domains, summarised in *Table 2*, where each question (excluding *living in own home*) has four responses. The meaning of these response options is summarised in *Box 1*.

In developing a new set of questions, we set out what each of the response options would look like for each of the domains. This is summarised in *Appendix 3*. This activity helped us to think about appropriate wording, and ensured that the whole team was clear about what each domain meant and what situation each response option was intended to represent. Where possible, we used terms such as 'want' or 'like' and 'adequate' to distinguish between the preferred situation and the 'needs met' or 'mustn't grumble' state, the aim being to reflect capability differences. Need states were described using terms that reflected amount or frequency of events, closer to functioning states. Once the items had been developed, we assessed them using the questionnaire appraisal system<sup>63</sup> before testing them with service users. We explain how the questions were tested, and present the results of this testing in the next section.

## Testing the revised Adult Social Care Outcome Toolkit instrument

### Method

The questions that form the basis of the current and expected SCRQoL scales, as well as the two methods for eliciting preferences (see *Chapter 4*), were tested and the wording developed so that it made sense to service users in cognitive interviews. Cognitive interviewing is a technique that has gained popularity over recent years as a way of exploring the validity of questions, in particular the aspect of standardisation in the meaning of the question across respondents.<sup>64,65</sup> The technique draws on cognitive psychology and separates the process of responding to the questions into four tasks or components: comprehension, retrieval (from memory using strategies), judgement and response (includes mapping the judgement to a response category and any editing of the response that may be required to present the respondent in a positive light).<sup>66</sup>

### BOX 1 Meaning of the options for the answers

*Ideal* The preferred situation, in which needs are met to the desired level

*No needs* Where needs are met, but not to the desired level

*Low-level needs* Where there are needs, but these do not have an immediate or longer-term health implication

*High-level needs* Where there are needs and these have an immediate or longer-term health implication

**TABLE 2** Domains of SCRQoL and definition of domain prior to testing with service users

Domain	Definition
Personal cleanliness and comfort	The service user feels that he/she is personally clean and comfortable and looks presentable or, at best, is dressed and groomed in a way that reflects his/her personal preferences
Accommodation cleanliness and comfort	The service user feels that his or her home environment, including all the rooms, is clean and comfortable
Food and drink	The service user feels that he/she has a nutritious, varied and culturally appropriate diet with enough food and drink that he/she enjoys at regular and timely intervals
Safety	The service user feels safe and secure. This means being free from fear of abuse, falling or other physical harm, and fear of being attacked or robbed
Social participation and involvement	The service user is content with his or her social situation, where social situation is taken to mean the sustenance of meaningful relationships with friends and family, and feeling involved or part of a community should this be important to him/her
Occupation	The service user is sufficiently occupied in a range of meaningful activities, whether formal employment, unpaid work, caring for others or leisure activities
Control over daily life	The service user can choose what to do and when to do it, having control over his/her daily life and activities
Dignity	The negative and positive psychological impact of support and care on the service user's personal sense of significance
Living in own home	The service user is living in his/her own home

These components are used by the interviewers as a guide during the interview process in order to uncover differences in the interpretation of questions by respondents.

Thirty cognitive interviews were undertaken during February, March and April 2009, with participants from a range of backgrounds, in three waves of 10 interviews across England. Service users were recruited through councils that had volunteered to act as 'critical friends' to the project. All participants had some contact with social care services, although some arranged their own support through a direct payment or personal budget. Most (20) were female and were white (29). About half of the sample were older (aged  $\geq 65$  years) – although eight were aged 40–64 years – and most lived alone. Only two were in full-time employment, although five were involved in unpaid work. They suffered from a range of conditions, including depression and anxiety, stroke, Parkinson's disease and arthritis.

After each wave of interviews, the three field researchers met to discuss the issues that arose for the interviewees as they tried to respond to the questions.<sup>67</sup> Discussion between the researchers centred on issues that allowed insight into the interviewees' thought processes as indicated by the cognitive model, and any problems in the wording were rectified and the questions retested in subsequent waves of interviews.

## Results

### Measurement domains

Identifying the aspects of QoL that matter most was an important part of developing the questionnaire. As we were trying to measure social care outcomes, the most important aspects were those that are affected by the provision of social care. The sensitivity of the measure depends on capturing the aspects that are affected by social care and matter to people. Everyone we interviewed agreed that the questions captured important aspects of their QoL.

### Wording of questions

We tested the wording of the questions and the answer options in order to come up with ways to express each aspect of SCRQoL in everyday language. We also wanted to find words for the answer options that captured the states in *Box 1*.

For *accommodation cleanliness and comfort*, we used the wording ‘my home is clean and comfortable’. Important aspects were having clean, dust-free surfaces and hygienic kitchens and bathrooms, but people also mentioned the state of the décor, whether or not their home was neat and tidy, whether or not their home had their own ‘stuff’ in it that they could get to easily, and also whether or not they could get around their home easily. For example, a number of people mentioned the problems posed by stairs, or how they had had their home adapted to make it easier for them to move around without help.

The term *control over daily life* was understood by the people we interviewed. They often distinguished between making decisions and carrying out those decisions. Most of the people we spoke to depended to some extent on help from others to see these decisions through. Having control over their daily life depended on them having someone – and, importantly, the right someone – to help them.

*Dignity* can mean many things to people, but ‘the way I think and feel about myself’ captured a person’s sense of self and significance well. Including ‘the way I’m helped and treated’ forced people to consider the way their care and support packages had an impact on their sense of self and significance. However, this was not the case for everyone. Some people felt negatively about themselves, not because of how they were treated but because it was difficult for them to accept that they needed help. Some of these people chose the ‘no needs’ answer, but others chose one of the ‘low- or high-level needs’ answers. In order to allow investigation of the impact of this on responses and interpretation of these, an additional question was asked about the impact of having help at all on how people felt about themselves.

We used the term ‘food and drink’ to express the aspects of meals and nutrition. Including ‘drink’ was very important, as people drink more often than they eat, and many people discussed how they managed their lives to ensure that they had the drink they needed. Initially the wording was ‘I can get ... the food and drink ...’, but we found that some people interpreted the ‘can’ too literally as being physically able to get food and drink without help. We changed the wording to ‘I get ... the food and drink ...’, which did not have the same problems.

‘Doing things I value and enjoy’ seemed to capture the type of things we intended for the *occupation* domain. People talked about voluntary work and paid work; activities they did with others, such as going out shopping or to eat; and activities they did on their own, such as reading, needlework or making cards. The answer options changed significantly from the first draft to capture the frequency of doing things, the quality of the things done, and the number of things people had to do. This is important, as for some people the problem was not that they did not have enough to do or things they enjoyed doing, but because they were not able to do as many things as they would like to do because of health limitations.

‘Clean and presentable’ was the term used for *personal cleanliness and comfort*, which was understood well by people. People talked about how frequently they washed, showered or bathed, and also about whether or not they were able to do their hair as they liked and wear the clothes they liked. Many of the women we interviewed talked about the difficulties they had with jewellery and make-up, and how important it was for them to be able to wear them.

*Feeling safe* was understood by everyone, although some people questioned whether we meant their feeling safe outside or only inside their home. We therefore recommended a prompt for interviewers to make it clear that we mean both outside and inside the home.

We had some difficulty finding a good way of expressing *social participation and involvement*, but settled on the phrase ‘social contact with people I like’. Similarly to the *occupation* aspect, it

is important to reflect the quality of contact, the frequency of contact and the number of people known to the person being interviewed. However, it was not possible to reflect all three aspects throughout the question, and the quality part – ‘people I like’ – was dropped from the last three options. This did not seem to matter: people continued to talk about contact with friends and family. They also mentioned phone, e-mail and letter contact, as well as face-to-face contact.

The original answer options also included the phrase ‘I feel lonely’, but this was taken out as it was confusing. As one person explained, ‘it depends on whether you mean personal or social life, as a person can be lonely because they don’t have a special person in their life or lonely because they don’t know many (or any) people’. Given the areas over which social care can be expected to have an impact, we felt it was important to focus the question on the social – rather than the personal – side. We chose to use the term ‘socially isolated’ to denote social loneliness, which seemed to work well.

We also tested including a time frame in the answer options of ‘the past couple of weeks’. We found, however, that it made the question difficult for people to follow and, because many people had conditions that fluctuated, they tended to ignore the instruction even when it was pointed out to them. Most people preferred to answer according to an ‘average’ day. We decided to leave these instructions out.

In the original instrument *living in own home* was also included. This was not asked as a question but was assessed by the interviewer according to where the person lived. It was, however, included in the preference elicitation tasks as an attribute. The team hypothesised that there may be well-being effects from living in your own home over and above those derived from the other aspects of SCRQoL. In the testing, when service users imagined not living in their own home, they tended to imagine living in a care home. When we asked them to describe what this was like, they described it in terms of the other aspects of QoL that were already included in the measure, such as *control over daily life, social participation and involvement*, and so on. This seemed to be evidence that this attribute was not distinct from the other aspects of SCRQoL, and therefore would not meet condition of preference independence.<sup>68</sup> We therefore decided to exclude this attribute from the measure.

### Response options

As well as understanding the domains as expected, people also seemed to interpret the response options as we had intended. However, for the *dignity* and *safety* questions, we felt that there was not enough difference in meaning between some of the levels (and these findings were confirmed by the preference elicitation feasibility study). For *dignity*, the problem was for the bottom two levels, which were changed from ‘sometimes undermines’ and ‘undermines’ to ‘sometimes undermines’ and ‘completely undermines’. The bottom two levels of *safety* were also seen as quite similar. These were changed to ‘feel less than safe’ and ‘don’t feel at all safe’ from ‘sometimes I don’t feel safe enough’ and ‘most of the time I don’t feel safe enough’. The top two levels of *safety* were also changed, with the second level amended to emphasise the sense of adequacy that is intended to be conveyed by this response option.

### The impact of services on quality of life: expected social care-related quality-of-life questions

As described in *Chapter 1*, we wanted to test a pragmatic approach to reflect the outcome of care services. In this, we asked people to assess, in addition to their current status, what their QoL within each domain would be like without services. In the cognitive testing, we tested these questions with service users. For the most part, people could visualise this ‘hypothetical’ situation in the absence of services, and their reasoning made sense and seemed plausible. Some examples of the reasoning of interviewees to these questions are given in *Box 2*.

**BOX 2** Examples of responses to the expected SCRQoL questions

A woman who had help with bathing and carrying washing to laundry was very worried about falling and a lack of bowel control. In the absence of services, she chose the high-level needs option for the *safety* domain, arguing that she would feel very unsafe bathing and walking about

A woman who had help with both shopping and cooking chose the high-level needs option for the *food and drink* domain in the absence of services, as she perceived that her health would be at risk as she is diabetic

A woman who had help with the *accommodation cleanliness and comfort* domain chose the high-level needs option in the absence of services, as she said she would be unable to do any housework because of her mobility problems. Interestingly, this issue was reflected in her responses to the *social participation* domain, where she chose high-level needs in the absence of services. She argued that, without services, she would not want anyone to come to her home, as it would be dirty from the lack of housework, and she would not be able to go to her bridge club as she needs someone to take her there

For the *personal cleanliness and comfort* domain, a woman who had help with personal care and washing clothes chose the high-level needs option in the absence of help from carers because she felt she would not be able to wash her back, take clothes out of the washing machine and hang them out

A woman who had help in the *personal cleanliness and comfort* domain chose the 'no needs' option in the absence of services (her current situation was the preferred situation). She argued that she would struggle to make herself as clean and presentable as she wished. She would not be able to shower, so would strip wash every time. She did add that this is what made her collapse before, as trying to manage these activities exhausts her. She ended up in hospital relatively recently and the doctors told her she needed to have a carer. Interestingly, this was reflected in the *safety* domain, where she said that in the absence of services she would have high-level needs as she would 'be in hospital more than out' because she would 'wear [herself] out trying to do things'

For the *social participation* domain, the same lady argued that, without her carers taking her out, as well as the contact she has with friends, she would not have the same kind of social life: people would come to her but she would not go out. For this reason she chose the low-level needs option in the absence of services

Despite evidence that these questions seemed to work on the whole, there were some examples where people responded in unexpected ways. In one case, a respondent chose low-level needs as the current SCRQoL state and 'needs met in the absence of services' arguing that 'they did not want lots of people coming in to provide help'. This response seems to be driven by a wish to send a message to those organising the care package, despite the fact that the interview findings were not shared with the relevant care manager.

It was also clear from the interviews that the task required a level of cognitive ability that some of the interviewees did not possess. As the interviews also included testing of the preference elicitation tasks, we were able to compare the cognitive ability required for each of the tasks (current SCRQoL questions, expected SCRQoL questions and preference elicitation study). One interviewee provided unclear reasoning for a number of the responses to the expected SCRQoL questions. For example, for the *personal cleanliness and comfort* domain, this person chose needs met in the absence of services arguing that, without help, it would not be possible to wash or shower but it would be possible to get dressed. This same person was incapable of understanding the preference elicitation task, and the interview had to be curtailed. It would seem that a similar level of cognitive ability is required for both the preference elicitation study and expected SCRQoL questions.

Another difficulty encountered was trying to get the interviewees to answer questions without imagining someone else stepping in, for instance a family member or private support. For example, a lady who had help from her carers to get and prepare food argued that in the absence of services she would have low-level needs, as she would find a way to manage even if it was



difficult for her on her own. She might ask someone to deliver food to her, so she would get enough food to sustain her, although it might not be the right food. When interviewers ask these questions, it is important to emphasise that interviewees should not try to imagine what other types of services or people could step in to help them out.

Another issue apparent from the cognitive interviews was that, prior to asking certain questions, it is important to clarify what services a person is receiving. Depending on the types of services the interviewees were receiving, we varied the services they should be thinking about in this testing phase. We asked people to include help from social services and specified exactly what services they should include, i.e. home care, day centres and so on. Where the interviewees did not receive help from social services, we asked them to think about any help they purchased privately. We also asked people to think about equipment they had, although this was more difficult for people to consider. What is included as 'services' and the precise wording will depend on the purpose of the study.

Finally, the importance of the specific words chosen was underlined by these interviews. People conceptualised 'support and services' very differently. For instance, few participants included equipment or adaptations when thinking about services. The researchers were able to identify such 'mistakes' only because they had asked each service user what services they were receiving before asking these questions. It is also important to tailor instructions to the service user. Thus, although the precise nature of wording will always depend on the nature of the study, in operationalising the measure we include detailed questions on service receipt, which can be used to tailor the wording to each individual's situation.

## Conclusions

The OSCA project built on previous versions of the ASCOT measure, drawing on conceptual developments, the literature, consultations with stakeholders and empirical evidence to cognitively test a revised measure with service users. This measure included a new domain (*dignity*) and a fourth level for all of the domains intended to increase sensitivity and reflect capabilities. The cognitive testing suggested that in taking the *dignity* domain forward it would be helpful to include an additional question that allowed respondents to reflect their general attitude to receiving any help, as some people have problems coming to terms with needing help at all. The *living in own home* domain was dropped as it appeared to be double-counting other domains, rather than reflecting an additional care setting or location dimension to our measure. The next step was to explore the validity of the revised measure.



## Chapter 3

# Testing validity

### Introduction

Validity assesses the extent to which an instrument measures what it is intended to represent. In Messick's words,<sup>69</sup> it is 'an integrated evaluative judgement of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of interpretations and actions based on test scores or other modes of assessment'. The ASCOT measure is intended to be of use in economic evaluations and to provide information for decisions about resource allocation across social care. It should enable decision-makers to compare the value of different types of social care provision, such as a meals service with a home care worker. Validating ASCOT is therefore about identifying the extent to which the instrument captures the value of social care.<sup>70</sup>

The ASCOT consists of several components to capture the value of social care. The main instrument is composed of two types of questions. The first set of questions asks people to rate their current SCRQoL state in terms of eight domains or attributes. A second set of questions then requires people to rate their current SCRQoL state in the absence of the 'intervention', within seven of the eight attributes, where 'intervention' can be defined variously according to the purpose of the study. The *dignity* attribute does not have an item for the second set of questions because it is process based and it cannot be asked when people are not receiving services. We refer to the second set of questions as the 'expected' SCRQoL items and the first set as the 'current' SCRQoL items. The third component of ASCOT is a set of preference weights that can be used to attach a value to each SCRQoL state. (The generation of preference weights is described in *Chapters 4 and 5*.)

The different sets of questions serve different purposes and can be used in a variety of contexts. The current items capture the prevailing SCRQoL state of the individual and could be used to compare the states of otherwise equivalent groups (such as matched samples). They could also be used for evaluation of interventions, where they could be administered before and after an intervention, to generate pre- and post-test scores, where the pre-test scores act as a proxy for the expected SCRQoL state in the absence of the intervention and the difference in the estimated effect. The alternative, pragmatic approach proposed in *Chapter 1* is to ask individuals what their expected SCRQoL in the absence of services is directly, with the difference between that and currently experienced SCRQoL representing the contribution that social care makes to a person's SCRQoL. For the expected score, *dignity* is assumed to be at the second level – where the care process has no impact on the person's sense of self-worth. All of the SCRQoL measures (current, expected and gain) can be preference weighted.

The psychometric criteria of 'construct under-representation' (the failure to capture important aspects of the concept being measured) and 'construct-irrelevant variation' (when responses to the measure are influenced by factors irrelevant to the concept being measured) are useful for thinking about validity in the context of valuing social care.<sup>71,72</sup> However, as Brazier *et al.*<sup>70</sup> recognise, the psychometric approaches used to determine validity need modification to make them applicable to a preference measure. Brazier *et al.*<sup>70</sup> identify three aspects of preference measures that require validation: the descriptive system, the valuations or utility weights, and the

empirical validity of the instrument, which refers to whether people, through their behaviour in practice, appear to value the different states in the way that they are valued in the measure.

Here we focus our assessment of validity on the validity of the descriptive system, which refers to the choice of domains, the specification of the items in the instrument, and the ability of the instrument to detect changes or known differences in SCRQoL. This was achieved in four separate sets of analysis. We first examined the construct validity of the individual items in terms of whether or not they reflect the concepts as intended. We then evaluated the construct validity of the three preference-weighted scales by exploring their ability to detect known differences in SCRQoL. In this we wanted to establish:

- current SCRQoL scale as a measure of social care-related QoL
- expected SCRQoL scale as a measure of social care need
- gain in SCRQoL scale as a measure of the contribution of services to SCRQoL.

## Methods

### Data collection

Throughout the project, in order to access service user samples, we made use of the annual UES conducted by local councils. The main data collection conducted to test the validity of the instrument with service users took place in 2009 when the UES was of older people (aged > 65 years) using home care services.<sup>13</sup> Ten councils across England took part, covering a variety of regions and local authority (LA) types: six shire counties, two London boroughs, one metropolitan district and one unitary authority. A sampling frame was generated from respondents who had indicated that they were happy to be approached to take part in further research. Data were collected face to face through computer-aided personal interviews (CAPIs). Interviewers were briefed prior to interviewing. Data collected included sociodemographic information; service receipt and informal support; QoL and psychological well-being; health; functional ability; control and autonomy; nature of the locality and environment; social contact and support; and participation in groups and volunteering.

### Analysis

The content of the instrument is clearly an important aspect of the validity of the descriptive system. If key aspects of SCRQoL relevant to a person's utility function are absent, the instrument will not provide an adequate valuation of social care. We followed the method used by Coast *et al.*,<sup>73</sup> who assessed validity by observing relationships between the items of their measure (ICECAP, now renamed ICECAP-O) and other factors thought to be related to it. Variables were divided into thematic groups for testing associations between these and items in the ASCOT measure. We examined the statistical significance of associations, and considered patterns of percentages and means to form a judgement about the strength or otherwise of relationships.

To demonstrate the validity of the current scale as a measure of SCRQoL, the aim was to explore its relationship with other variables that capture the same construct. However, because of the uniqueness of this measure – in its focus on SCRQoL – it was difficult to find measures with which to compare its performance. We therefore examined its relationship with other measures capturing related constructs. These included HRQoL, where we would expect a moderate relationship with SCRQoL, and psychological well-being, where we might expect a closer relationship.

To reflect HRQoL we used the European Quality of Life-5 Dimensions (EQ-5D),<sup>74,75</sup> a widely used indicator that has preference weights that generate a measure of health value. For psychological well-being we used the 12-item version of the General Health Questionnaire (GHQ-12). Although originally developed as a measure of mental ill health, with a cut-off score below which it is likely the person is clinically depressed,<sup>76–78</sup> GHQ-12 has been tested as a measure of positive mental health in the general population.<sup>79</sup> In addition, as ASCOT attempts to capture capability, we anticipated a moderate relationship with measures of concepts such as control, autonomy and independence. To reflect these concepts, we used the control and autonomy subscale of the CASP-12,<sup>80</sup> a reduced form of CASP-19, which is a theoretically based needs satisfaction measure of quality of life for older people.<sup>81</sup> The items capturing expected SCRQoL in the absence of services can be viewed as measuring the need for social care services, as the items capture what a person's life would be like without the compensatory action of services. We would therefore expect the expected SCRQoL scale to be associated with other measures that capture need for help in activities of daily living (ADL) and instrumental activities of daily living (IADL). These measures capture the functional ability and are frequently used in needs assessments for social care.

The SCRQoL gain measure is designed to capture the contribution of services to SCRQoL, so construct validity was explored in relation to service receipt. 'Services' here were any publicly funded service and included home care, day centres and meals services, as well as newer forms of service delivery, such as direct payments. We would expect the gain in SCRQoL measure to have a positive correlation with intensity of service receipt, although the strength of the correlation will depend on a number of factors that affect the production of welfare, including the quality of the care delivered and other factors that may influence the ability of workers to deliver optimal care, such as the design of the person's home or challenging behaviour of the individual. It is also possible that the relationship is non-linear, as increasing levels of service input deliver diminishing marginal returns. Therefore, we would not necessarily expect a strong relationship, but merely that the correlation is significant and positive.

We examined the relationships with our individual SCRQoL items using chi-squared tests (for unordered or ordered categorical variables) or one-way analysis of variance (for continuous variables). For comparisons with the SCRQoL current, expected and gain scales, we used a series of Pearson correlations with continuous variables and one-way analysis of variance (for unordered or ordered categorical variables).

## Description of sample

A total of 566 contacts were attempted from a sample of 778 respondents to the older home care UES who were willing to participate in further research; this produced 301 (53%) complete interviews. A few refused ( $n = 18$ , 3%) or had died ( $n = 4$ , 1%), but most non-participants were not contactable ( $n = 243$ , 43%). The characteristics of the sample are shown in *Table 3*. As would be expected among publicly funded older home care service users, the majority of the sample was female, > 80 years old, single and living alone.<sup>13,82</sup> However, compared with the population of publicly funded older home care users from which this sample was drawn, the sample had slightly fewer females (68% compared with 72%), slightly more people from a white ethnic background (98% compared with 93%) and a smaller proportion of people > 80 years old (60% compared with 68%).<sup>13</sup>

**TABLE 3** Sociodemographic and economic characteristics of sample members

Characteristic	Frequency	%
<b>Sex (n = 301)</b>		
Female	205	68.1
<b>Ethnicity (n = 296)</b>		
White	296	98.3
<b>Age (years) (n = 301)</b>		
65–69	27	9.0
70–79	95	31.6
80–89	137	45.5
≥ 90	42	14.0
<b>Area of residence</b>		
London borough	35	11.6
Another large city or town	59	19.6
Suburb of large city/town	47	15.6
Small town	98	32.6
Rural area or village	62	20.6
<b>Marital status (n = 301)</b>		
Married/living together	82	27.2
Never married	30	10.0
Widowed	168	55.8
Separated/divorced	19	6.3
<b>Living situation (n = 301)</b>		
Live alone	202	67.1
<b>Tenure (n=301)</b>		
Owner–occupier	154	51.2
Rent-paying tenant	142	47.2
Tenant living rent-free	5	1.7
<b>Income (£, per week) (n = 182)</b>		
≤ 275	122	40.5
276–374	40	13.3
375–424	13	4.3
425–574	2	0.7
≥ 575	5	1.7

## Current social care-related quality of life

### Individual domains

The distribution of responses to the current SCRQoL items is shown in *Table 4*. All questions were answered by all respondents (a much improved response rate compared with that found in previous versions of the measure – see *Chapter 2*). For most of the items, there was a good spread of responses across all levels of the attributes, although for all of the items, except the *occupation* item, the distribution is skewed towards the more positive end of the response scale. If services are effective then a skewed distribution is inevitable with a measure such as this one: the top

**TABLE 4** Distribution of responses to the current SCRQoL items

Item	Frequency	%
<b><i>Control over daily life</i></b>		
I have as much control over my daily life as I want	89	29.6
I have adequate control over my daily life	92	30.6
I have some control over my daily life	94	31.2
I have no control over my daily life	26	8.6
<b><i>Personal cleanliness and comfort</i></b>		
I feel clean and am able to present myself the way I like	176	58.5
I feel adequately clean and presentable	102	33.9
I feel less than adequately clean or presentable	21	7.0
I don't feel at all clean or presentable	2	1
<b><i>Food and drink</i></b>		
I get all the food and drink I like when I want	238	79.1
I get food and drink adequate for my needs	60	19.9
I don't get all the food and drink I need, but no health risk	1	0.3
I don't get all the food and drink I need, and health risk	2	0.7
<b><i>Accommodation cleanliness and comfort</i></b>		
My home is as clean and comfortable as I want	193	64.1
My home is adequately clean and comfortable	100	33.2
My home is less than adequately clean and comfortable	7	2.3
My home is not at all clean and comfortable	1	0.3
<b><i>Personal safety</i></b>		
I feel as safe as I want	169	56.1
I feel adequately safe, but not as safe as I'd like	94	31.2
I feel less than adequately safe	30	10.0
I don't feel at all safe	8	2.7
<b><i>Social participation</i></b>		
I have as much social contact as I want with people I like	110	36.5
I have adequate social contact	99	32.9
I have some social contact with people, but not enough	67	22.3
I have little social contact with people and feel socially isolated	25	8.3
<b><i>Occupation</i></b>		
I'm able to spend my time as I want, doing things I value or enjoy	86	28.6
I'm able to do enough of the things I value or enjoy with my time	77	25.6
I do some of the things I value/enjoy with my time, but not enough	115	38.2
I don't do anything I value or enjoy with my time	23	7.6
<b><i>Dignity</i></b>		
The way I'm helped and treated makes me think and feel better about myself	150	49.8
The way I'm helped and treated does not affect the way I think/feel ...	115	38.2
The way I'm helped and treated sometimes undermines the way I think/feel ...	33	11.0
The way I'm helped and treated completely undermines the way I think/feel ...	3	1.0

two levels are designed to reflect the state of no needs and the bottom two levels are designed to reflect the states of low- and high-level needs, respectively, so where services are effective we should find very few people with needs across any of these items. This is particularly likely to be the case for the basic domains (*personal care, safety, food and drink, and accommodation cleanliness and comfort*), as compensating people for needs in these areas is arguably the primary purpose of social care services. However, the distribution of responses to the *food and drink* and *accommodation cleanliness and comfort* items was concerning, and was out of step with the *personal care* and *safety* items. Comparison of these responses with distributions of the three-level version of the items that formed part of the UES form through which the respondents were recruited indicated that the poor distributions may have been a result of the wording, and in particular the wording of the new 'no needs' levels.

The content of the instrument is clearly an important aspect of the validity of the descriptive system. The comparison of the individual items with variables reflecting themes hypothesised to be associated with each SCRQoL domain is reported in detail elsewhere.<sup>83</sup> In general, where we anticipated relationships, these were found, with strong relationships between our measure and indicators of well-being (as measured by GHQ-12) and control and autonomy (as measured by the subscale of CASP-12), and the attributes as we would hope (*Table 5*). A few findings were unanticipated, but these could be plausibly explained. For example, we had expected that *social participation* and *occupation* would be associated with the type of area people were living in. It was, but poorer *social participation* was experienced by those living in suburban areas and small towns rather than, as we had anticipated, those living in rural areas. We concluded that this might be related to council policies in the areas in our sample or due to suburban areas lacking a sense of community that is strong in rural areas.<sup>84</sup>

The poor distribution of responses to the *accommodation* and *food and drink* items, together with parallel findings with respect to lack of distinction by respondents in the MOPSU preference study<sup>10</sup> of this version of the instrument, led to the decision to revise the wording for these items. The amended versions were cognitively tested and piloted through parallel work with service users, which was being conducted to feed into the national ASCS (see *Table 1*).<sup>8</sup> In order to ensure that the final items used in our measure were performing well, we repeated the analysis above using the sample of 458 users of equipment services who participated in the preference study reported in *Chapter 5*. Not all of the variables we used were available, but we were able to use substitutes for most themes. The results are presented in *Appendix 4*. The distribution of the final items for that sample is shown in *Chapter 6*.

### Summary score

Each domain was scored '0' (high needs) through to '3' (ideal state) and the scores for each domain were summed for each individual to produce a single score of current SCRQoL, taking values from 0 to 24. The mean score for this sample was 18.2 [standard deviation (SD) = 3.49, maximum (max) = 24, minimum (min) = 8]. The domain scores were also preference weighted using the provisional valuations from the MOPSU preference elicitation study.<sup>10</sup> The SCRQoL score based on this could take values from 8.13 to 35.97. The mean score for this sample was 30.3 (SD = 4.1, max = 35.97, min = 19.04) (*Figure 1*). The distribution of the two scores was similar and there were very few, very minor differences between the results reported here for the equal- and preference-weighted measures, so we have reported only results for the preference-weighted scores.

*Table 6* shows that current SCRQoL was significantly correlated with the measures of well-being, control and autonomy, and HRQoL. The relationship with the CASP-12 control and autonomy subscale and the GHQ-12 was strong, with correlation coefficients being >0.5 and close to 0.6.<sup>85</sup> As expected, there was a significant relationship with our indicator of HRQoL, EQ-5D, but less strong.



**TABLE 5** Mean GHQ-12 and CASP autonomy and control scores by current SCRQoL attribute

Attribute	n	GHQ-12	CASP
<b>Control over daily life</b>			
I have as much control over my daily life as I want	89	25.97 <sup>a</sup>	11.96 <sup>a</sup>
I have adequate control over my daily life	92	23.68 <sup>b</sup>	9.90 <sup>b</sup>
I have some control over my daily life	94	20.83 <sup>c</sup>	8.22 <sup>c</sup>
I have no control over my daily life	26	16.62 <sup>d</sup>	6.89 <sup>c</sup>
Significance		<b>p &lt; 0.001</b>	<b>p &lt; 0.001</b>
<b>Personal cleanliness and comfort</b>			
I feel clean and am able to present myself the way I like	176	24.61 <sup>a</sup>	10.80 <sup>a</sup>
I feel adequately clean and presentable	102	21.08 <sup>b</sup>	8.49 <sup>b</sup>
I feel less than adequately clean and presentable	21	18.24 <sup>b</sup>	7.05 <sup>b</sup>
I don't feel at all clean or presentable	2	7.50 <sup>b</sup>	6.00 <sup>a,b</sup>
Significance		<b>p &lt; 0.001</b>	<b>p &lt; 0.001</b>
<b>Food and drink</b>			
I get all the food and drink I like when I want	238	23.74 <sup>a</sup>	10.00 <sup>a</sup>
I get food and drink adequate for my needs	60	19.52 <sup>b</sup>	8.67 <sup>b</sup>
I don't get all the food and drink I need (collapsed bottom two)	3	20.00 <sup>a,b</sup>	8.67 <sup>a,b</sup>
Significance		<b>p &lt; 0.001</b>	<b>p &lt; 0.05</b>
<b>Accommodation cleanliness and comfort</b>			
My home is as clean and comfortable as I want	193	24.24 <sup>a</sup>	10.41 <sup>a</sup>
My home is adequately clean and comfortable	100	20.60 <sup>b</sup>	8.56 <sup>b</sup>
My home is less than adequately clean and comfortable/My home is not at all clean and comfortable (collapsed)	8	17.63 <sup>b</sup>	7.75 <sup>a,b</sup>
Significance		<b>p &lt; 0.001</b>	<b>p &lt; 0.001</b>
<b>Safety</b>			
I feel as safe as I want	169	24.41 <sup>a</sup>	10.51 <sup>a</sup>
I feel adequately safe, but not as safe as I'd like	94	22.52 <sup>a</sup>	9.07 <sup>b</sup>
I feel less than adequately safe	30	17.53 <sup>b</sup>	7.67 <sup>b</sup>
I don't feel at all safe	8	13.88 <sup>a,b</sup>	8.50 <sup>a,b</sup>
Significance		<b>p &lt; 0.001</b>	<b>p &lt; 0.001</b>
<b>Social participation</b>			
I have as much social contact as I want with people I like	110	25.23 <sup>a</sup>	11.59 <sup>a</sup>
I have adequate social contact	99	22.99 <sup>b</sup>	9.40 <sup>b</sup>
I have some social contact with people, but not enough	67	19.94 <sup>c</sup>	8.12 <sup>b,c</sup>
I have little social contact with people and feel socially isolated	25	19.72 <sup>b,c</sup>	7.08 <sup>c</sup>
Significance		<b>p &lt; 0.001</b>	<b>p &lt; 0.001</b>
<b>Occupation</b>			
I'm able to spend my time as I want, doing things I value or enjoy	86	26.59 <sup>a</sup>	12.37 <sup>a</sup>
I'm able to do enough of the things I value/enjoy with my time	77	23.62 <sup>b</sup>	10.17 <sup>b</sup>
I do some of the things I value/enjoy with my time, but not enough	115	20.40 <sup>c</sup>	7.97 <sup>c</sup>
I don't do anything I value or enjoy with my time	23	18.61 <sup>c</sup>	7.13 <sup>c</sup>
Significance		<b>p &lt; 0.001</b>	<b>p &lt; 0.001</b>

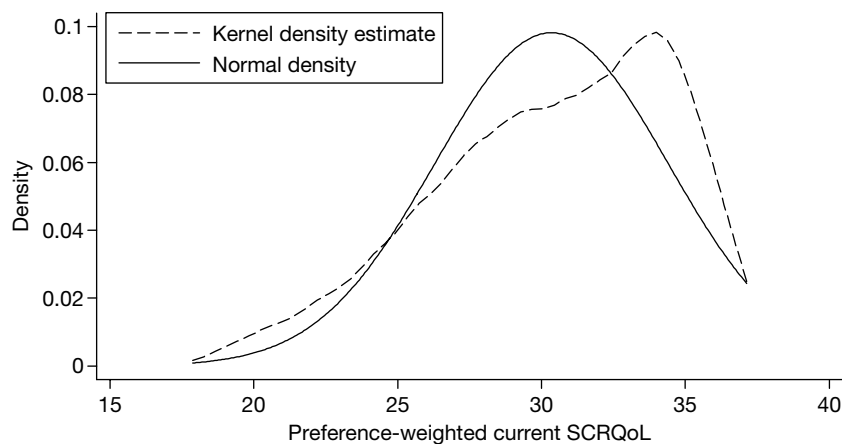
*continued*

**TABLE 5** Mean GHQ-12 and CASP autonomy and control scores by current SCRQoL attribute (*continued*)

Attribute	<i>n</i>	GHQ-12	CASP
<b>Dignity</b>			
The way I'm helped and treated makes me think and feel better about myself	150	23.48 <sup>a</sup>	9.73 <sup>a,b</sup>
The way I'm helped and treated does not affect the way I think/feel about myself	115	23.32 <sup>a</sup>	10.23 <sup>a</sup>
The way I'm helped and treated sometimes undermines the way I think/feel ...	33	19.00 <sup>b</sup>	8.06 <sup>b</sup>
The way I'm helped and treated completely undermines the way I think/feel ...	3	16.33 <sup>a,b</sup>	8.33 <sup>a,b</sup>
Significance		<b><i>p</i> = 0.001</b>	<b><i>p</i> &lt; 0.05</b>

a–d Means with different letter suffixes are significantly different from one another, as measured by post hoc tests (Tamhanes – equal variances not assumed).

High scores on GHQ-12 indicate poor psychological well-being so we would expect a negative sign for the correlation coefficient.



**FIGURE 1** Kernel density plot for distribution of scores to current SCRQoL measure. Kernel = Epanechnikov; bandwidth = 1.1678.

### Expected social care-related quality of life

The distribution of responses to the expected SCRQoL items is shown in *Table 7*. A small proportion of responses are missing for each item – between 1% and 4% of responses. The largest proportion missing is for the *safety* attribute, which implies that this is the hardest attribute for people to judge what their outcome state would be like in the absence of services. However, the missing responses were small for all items, indicating that this method is feasible with the population.

Unlike the responses to the current SCRQoL items shown in *Table 4*, where the effect of services is captured in the SCRQoL state, in this measure of expected SCRQoL responses are much more evenly spread across all the response options, reflecting the fact that for this sample their SCRQoL would be much poorer in the absence of services. Only the *food and drink* attribute had over half the sample in the top 'no needs' level, although the *accommodation cleanliness and comfort* attribute also approached 5%, again suggesting that the wording was not working as we might have hoped.

As with current SCRQoL, the expected SCRQoL values can be summed to produce an equally weighted score of expected SCRQoL, which can take values from 0 to 24. The mean score for this sample is 13.8 (SD = 4.40, max = 23, min = 2, *n* = 281). Using the valuations from the

**TABLE 6** Correlation between preference-weighted current SCRQoL scales and other measures of aspects of QoL ( $n=301$ )

Measures	Mean score (SD)	Correlation with SCRQoL
EQ-5D	0.29 (0.36)	0.41***
GHQ-12	13.14 (6.50)	-0.58***
Control and autonomy subscale of CASP-12	9.72 (3.67)	0.58***

\*Significant at 10% level; \*\*significant at 5% level; \*\*\*significant at 1% level.

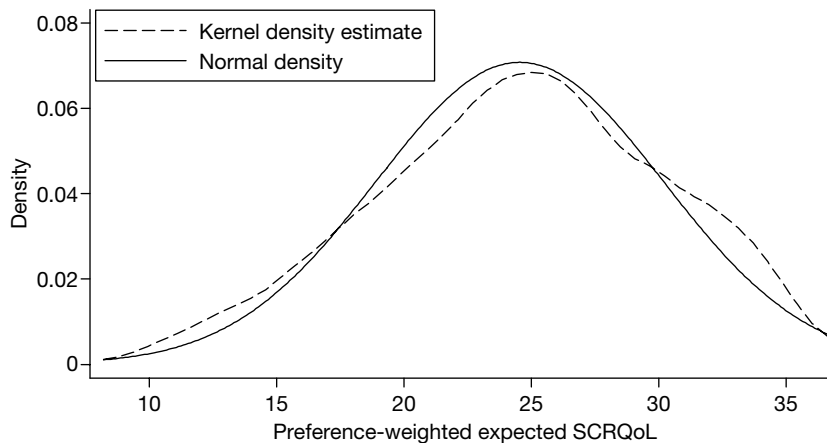
**TABLE 7** Distribution of responses to the expected SCRQoL items

Attribute	Frequency	%
<b>Control over daily life</b>		
I would have as much control over my daily life as I want	29	9.8
I would have adequate control over my daily life	41	13.9
I would have some control over my daily life	123	41.6
I would have no control over my daily life	103	34.8
Total	296	100.0
<b>Personal cleanliness and comfort</b>		
I would feel clean and able to present myself the way I like	85	28.6
I would feel adequately clean and presentable	62	20.9
I would feel less than adequately clean or presentable	71	23.9
I would not feel at all clean or presentable	79	26.6
Total	297	100.0
<b>Food and drink</b>		
I would get all the food and drink I like when I want	160	54.1
I would get food and drink adequate for my needs	71	24.0
I wouldn't get all the food and drink I need, but no health risk	18	6.1
I wouldn't get all the food and drink I need, and health risk	47	15.9
Total	296	100.0
<b>Accommodation cleanliness and comfort</b>		
My home would be as clean and comfortable as I want	144	48.3
My home would be adequately clean and comfortable	84	28.2
My home would be less than adequately clean and comfortable	39	13.1
My home wouldn't be at all clean and comfortable	31	10.4
Total	298	100.0
<b>Personal safety</b>		
I would feel as safe as I want	100	34.6
I would feel adequately safe, but not as safe as I'd like	73	25.3
I would feel less than adequately safe	61	21.1
I wouldn't feel at all safe	55	19.0
Total	289	100.0

*continued*

**TABLE 7** Distribution of responses to the expected SCRQoL items (*continued*)

Attribute	Frequency	%
<b>Social participation</b>		
I would have as much social contact as I want with people I like	81	27.6
I would have adequate social contact	82	27.9
I would have some social contact with people, but not enough	85	28.9
I would have little social contact with people and feel socially isolated	46	15.7
Total	294	100.0
<b>Occupation</b>		
I would be able to spend my time as I want, doing things I value or enjoy	77	25.9
I would be able to do enough of the things I value or enjoy with my time	62	20.9
I would do some of the things I value/enjoy with my time, but not enough	116	39.1
I wouldn't do anything I value or enjoy with my time	42	14.1
Total	297	100.0

**FIGURE 2** Kernel density plot for distribution of scores to expected SCRQoL measure. Kernel=Epanechnikov; bandwidth = 1.6419.

MOPSU preference elicitation study,<sup>10</sup> the mean preference-weighted score for this sample is 24.5 (SD = 5.6, max = 34.97, min = 9.85,  $n = 281$ ). *Figure 2* shows the distribution of the preference-weighted scores. Both this and the equally weighted distribution were close to normal.

Activities of daily living were measured using a summary score of nine items: ability to manage stairs, go outdoors and walk down the road, get around indoors, get in and out of bed, use WC or toilet, wash face and hands, bath or shower or wash all over, get dressed and undressed, and feed oneself. Each ADL is scored on a scale from 0 to 3, where '0' = 'cannot manage at all', '1' = 'manage with help from someone else', '2' = 'manage on own with difficulty', and '3' = 'manage on own'. The correlation between expected SCRQoL and a summary score based on these nine ADLs was highly significant ( $r = 0.36$ ,  $n = 244$ ,  $p < 0.001$ ). This was higher than the correlation between current SCRQoL and ADLs ( $r = 0.29$ ,  $n = 258$ ,  $p < 0.001$ ), but not by much. Such measures are expected to be associated with current SCRQoL, as it is harder for people to achieve good SCRQoL because of the increasing difficulty associated with compensating people for the effects of severe impairments.

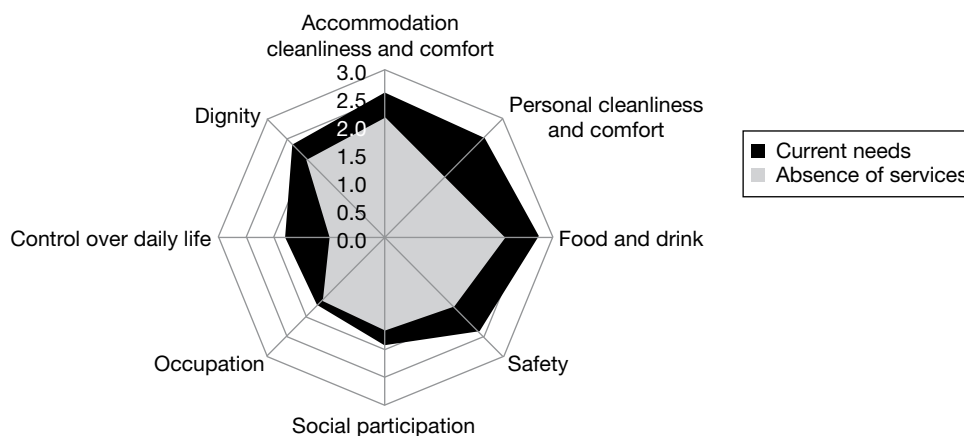
Social care-related quality of life in the absence of services is difficult to explain using only an indicator of functional ability, because of the complexity of need and the role of forms of support and context (such as home design) that may also contribute to SCRQoL in the absence of services. The primary objective in measuring expected SCRQoL is to provide a pragmatic approach to assessing the impact of social care interventions, by comparing this with people's current experiences, to measure SCRQoL gain.

## Social care-related quality-of-life gain

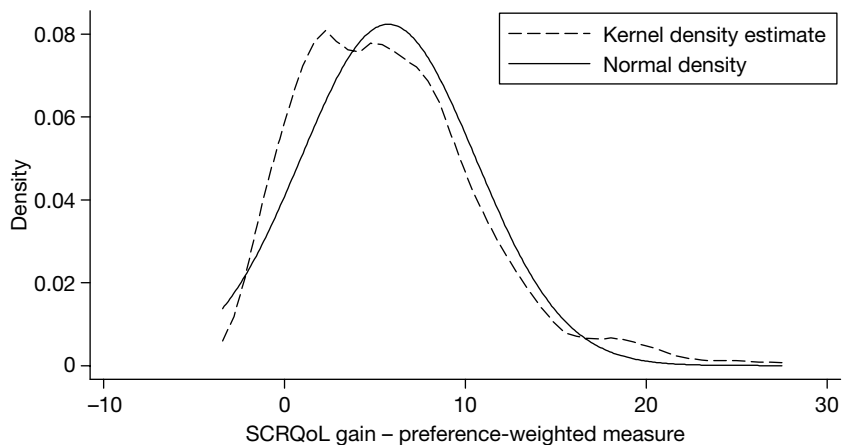
We can illustrate the impact of services on different domains using a cobweb chart. *Figure 3* shows this for this sample of older home care service users, with the darker-shaded area representing the impact of home care user service packages on SCRQoL. This shows that the contribution of services received by this group was greatest for *personal cleanliness and comfort* then *control over daily life* and smallest for *occupation* and *social participation*.

For each individual, the welfare gain (or loss) for each attribute can be summed together to produce a single score of impact of services on SCRQoL (SCRQoL gain), which can take values from  $-24$  to  $24$ , where negative values indicate a loss of welfare and positive values indicate a gain. (The value for expected *dignity* is given by assuming that in the absence of services 'The way I'm helped and treated does not affect the way I think/feel about myself' is chosen.) The mean score for this sample was  $4.42$  ( $SD = 3.77$ ,  $max = 22$ ,  $min = -3$ ). The two preference-weighted versions of this measure can also be subtracted from each other to generate a preference-weighted version of the gain in SCRQoL, which can take values from  $-35.97$  to  $35.97$ . The mean score for this sample is  $5.74$  ( $SD = 4.85$ ,  $max = 26.12$ ,  $min = -2.03$ ,  $n = 281$ ). The distribution of preference-weighted measure is skewed (*Figure 4*). The positive value for the sample means that, on average, this sample gained in SCRQoL as a result of social care. The distribution for the preference-weighted version was closer to normal.

The majority of the sample was using publicly funded home care services. A straightforward measure of service intensity is the number of hours of publicly funded home care received. However, 54 people in the sample reported not receiving publicly funded home care and, of those receiving publicly funded home care, a number also reported receiving other publicly funded services. Ideally, a measure of service intensity would capture these other sources of support. Therefore, we have constructed a second measure of service intensity: the cost of the care package, based on weekly levels of home care, day care and meals receipt, and unit costs of these for each council in our survey.<sup>13</sup>



**FIGURE 3** The contribution of services to SCRQoL for each attribute.



**FIGURE 4** Kernel density plot for distribution of scores to gain in SCRQoL measure. Kernel=Epanechnikov; bandwidth=1.3933.

Table 8 shows the correlation coefficient for the relationship between both of the service intensity variables and the gain in SCRQoL and current SCRQoL scales. We have also included the correlation between the log of service intensity and both versions of SCRQoL to account for the fact that the relationship is likely to be non-linear.

Interestingly, and consistent with research elsewhere using different measures of social care outcome, the SCRQoL current measure has a non-significant and, in many cases, a slightly negative correlation with service intensity.<sup>86,87</sup> We would expect people with greater needs generally to receive larger packages of care. No significant correlation would indicate that a similar SCRQoL is being achieved irrespective of need, suggesting that if services are being allocated rationally they are compensating as we would expect. Thus, current SCRQoL captures the result of care and other factors that influence SCRQoL.

The significant positive correlation between service intensity and gain in SCRQoL indicates that the more intense the service, the greater the gain in SCRQoL. There was some evidence that the relationship between service intensity and SCRQoL gain is non-linear, where service intensity is measured using hours of home care. Of course, in addition to the factors that we identified above as likely to be associated with expected SCRQoL, there will be other factors such as the quality of care provided. Multivariate analyses are required to establish the underlying relationship between service intensity and impact. Nevertheless, we have the expected relationship between service receipt and SCRQoL, suggesting that the gain in SCRQoL measure is performing as expected.

## Conclusion

The aim of the stages of work in the development of the measure reported in this chapter was to ensure that the measure to be preference weighted was practical and had good psychometric properties. The analyses conducted provide some evidence for the validity of ASCOT and the feasibility of using the measure with social care service users. Response rates were good for all of the items in both sets of questions, which implies that both the current SCRQoL items and the hypothetical pre-test approach used here are feasible with this population.

The detailed investigation into the validity of the measures also provided some evidence for validity. There was good evidence for the validity of the descriptive system for ASCOT: each attribute seemed to capture a different aspect of SCRQoL, and each attribute had the anticipated

**TABLE 8** Correlation between measures of service intensity and the current SCRQoL and gain in SCRQoL measures

Measure of service contribution	Correlation with current SCRQoL	Correlation with SCRQoL gain
Hours of publicly funded home care	-0.092 ( <i>n</i> =247)	0.232*** ( <i>n</i> =233)
Size of publicly funded care package	-0.084 ( <i>n</i> =271)	0.209*** ( <i>n</i> =254)
Log of hours of publicly funded home care	-0.103 ( <i>n</i> =247)	0.240*** ( <i>n</i> =233)
Log of size of publicly funded care package	-0.096 ( <i>n</i> =271)	0.184*** ( <i>n</i> =254)

\*Significant at 10% level; \*\*significant at 5% level; \*\*\*significant at 1% level.

relationship with other variables capturing similar and dissimilar concepts. There was also some evidence for the validity of the current and impact SCRQoL scales, which appeared to operate as intended.

Concern about the distributions of both the current and expected *accommodation* and *food and drink* items was compounded by a lack of differentiation in preferences for those items in parallel work that was undertaken on establishing preference weights for this version of the measure as part of the MOPSU project (see *Appendix 6*). This suggested that respondents did not see the 'capability' difference we were attempting to distinguish. Moreover, such weights combined with the distributions meant the items would not add substantially to the measure. We revisited the wording and took opportunities through parallel studies to test out the revised items with service users.<sup>8</sup> The final version of the instrument is shown in *Appendix 5*, and population and equipment service user distributions of the 'current' items using this are shown in *Chapter 6*.





## Chapter 4

# Preference elicitation development of methods

### Introduction

As described in *Chapter 1*, the first phase of the study included an investigation into alternative approaches for modelling preferences of the general population for social care outcomes, and an exploration of the applicability of these approaches to a service user population. The aim was to support the design of a large-scale population and service user preference study to develop weighted measures of social care. In addition, we wanted to explore the feasibility of anchoring the measure to the state of 'being dead', to provide the potential for using the measure as a social care equivalent of the widely used QALY in health care. The work was undertaken in parallel to the development work on the measure itself, so used early versions of the domains or attributes.

In this chapter we start with a brief consideration of the potential options open to us, before describing the results of the pilot feasibility study that investigated the use of DCE and BWS methods. We then describe the results of cognitive testing of these techniques with service users and investigation of the TTO method in the context of social care.

### Eliciting preferences

In eliciting preferences, there are a number of considerations that feed into the validity and reliability of the resulting preference weights. These include ensuring that the task is not too complex, so we can have confidence in the data collected from respondents. This is particularly an issue with our measure, which has a relatively large number of domains or attributes. Ideally, we also want to minimise the assumptions we make about what people are doing in expressing their preferences, in particular so that their responses can be interpreted on a cardinal basis. We want the process used to be consistent with expressions and interpretation of preferences in economic theory. In addition, ideally we want it to be possible to anchor the weights to some other characteristic (such as the 'being dead' state) or widely understood unit (such as money) in a way that provides potential for a wider use than a free-floating index that can only be compared with itself.

Typically, public preferences are elicited in nationally representative valuation exercises using the visual analogue (VA), standard gamble (SG) or TTO.<sup>68</sup> These tools require respondents to manipulate probabilities or lengths of life and so rely on an assumption of cardinality in responses. Theoretical and empirical problems with these methods<sup>88</sup> have been accompanied by interest in tasks that require only ordinality in responses, such as DCE and BWS.

Unlike the related field of conjoint analysis (CA), DCE and BWS are based on a sound, well-tested theory of decision-making and choice behaviour known as random utility theory (RUT).<sup>88-91</sup> CA is used to establish the relative importance of different attributes in the provision of a good or service, as well as to estimate the overall utility of a given commodity, with levels of given attributes. Individuals' preferences are recorded as rankings or ratings, both of which

are cognitively demanding tasks. The models developed (analysis of variance, ordered probit or ordered logit) are purely statistical, with no behavioural basis. Traditional CA investigates the process by which people evaluate one option at a time and state a preference for each option, which contradicts with most real-life decisions that involve comparisons of alternatives.<sup>92</sup> Experimental design techniques are used to construct a sample of scenarios – presented as combinations of attribute levels – from a full factorial set. In a DCE, respondents are asked to choose an option (scenario) out of a set of two or more options, which ultimately involves a comparison of indirect utility functions. Analysis of data is conducted using discrete choice models,<sup>92</sup> and, provided that any variance heterogeneity on the latent utility scale is properly accounted for,<sup>93,94</sup> the estimated coefficients of each of the attributes in the model are unbiased utility weights that represent the relative importance of domain levels.

Discrete choice experiments have been used extensively to facilitate analyses in the fields of transport and environmental policy. However, they can also be used to develop preference weights for different instruments, and work is under way to do so for the EQ-5D-5L (EQ-5D 5-level) for measuring health outcomes (as a supplement to a TTO valuation)<sup>95</sup> and the ICECAP capability indices.<sup>73,96</sup> They have also been used previously to explore the potential for developing preference-based outcome measures in social care,<sup>97,98</sup> including addressing the problem of a large number of attributes by using two overlapping experiments.<sup>68</sup>

The BWS method is an alternative preference elicitation method that also requires only an assumption of ordinality. It was developed by Louviere and Woodworth<sup>99</sup> and its first application was published in 1992<sup>100</sup> illustrating Case 1 (the ‘object’ case). The method gained popularity in health and social care when the properties of Case 2 (the ‘profile’ – previously called ‘attribute’ case) were proved,<sup>101</sup> and a guide to its use was published.<sup>102</sup> Flynn<sup>103</sup> provides an overview and theoretical discussion of the different cases of BWS. Case 2 has particular advantages in valuation studies that seek to elicit general population preferences for important attributes of QoL (or whatever maximand is of relevance to policy-makers<sup>104</sup>). In particular, it presents profiles one at a time, rather than in choice sets of size two or more as in a traditional DCE. This is important when respondents do not have experience of making choices in the particular area of application: keeping two or more profiles in mind at once is likely to be a harder task, leading to an increase the size of the random utility component and reduction of the statistical efficiency of the preference elicitation.

Best-worst scaling is a relatively new method, and there is a lack of empirical evidence that demonstrates the superiority of BWS over DCE in terms of practical considerations in the field (e.g. respondents’ fatigue) and the robustness of the modelling estimates. As shown in *Table 9*, DCE and BWS are promising approaches as they satisfy a number of theoretical and practical considerations. However, with such approaches there is an issue of anchoring. Anchoring refers to the adjustment of utility weights obtained from DCE or BWS experiments on to a ‘0–1’ scale, where ‘0’ represents ‘being dead’ and ‘1’ is the social care equivalent of ‘perfect health’, respectively. Anchoring also allows integration of information about the QoL and the length of time over which that QoL is maintained, and wider comparisons across interventions. However, this presents methodological challenges, both in terms of how to get respondents to consider a ‘being dead’ situation in the context of choices regarding the outcomes of social care, and how respondents interpret whether any given intervention (and resulting life state) is likely to increase or reduce the time until death. Moreover, including ‘being dead’ as a state (like the social care states) will neither conceptually nor empirically estimate the trade-off between length of life and QoL.<sup>104</sup>

One approach to the problem is to undertake a complementary piece of work to allow us to anchor the measure to the ‘being dead’ state, while using the main preference study to estimate

the individual preference weights. We identified TTO, seen as a simpler alternative to SG (as shown in Dolan *et al.*<sup>75</sup>) as a method that would allow us to explicitly identify the relationship between population preferences for SCRQoL states and ‘being dead’, thus, providing an anchor for the measure. TTO was chosen in preference to the other widely used cardinal technique of visual analogue scaling (VAS). TTO asks the respondent to explicitly consider a trade-off between length of life and QoL that has been argued to have a stronger grounding in theory than a simple rating task.<sup>103</sup>

The main tasks of the feasibility stage of this aspect of the work were, therefore, to compare the discrete choice and profile-case BWS to elicit values for different dimensions of SCRQoL. This involved evaluating both the extent to which valuations of SCRQoL states obtained through the two approaches produced similar results, and the relative complexity of the tasks, particularly from the perspective of service users. In addition, we needed to test how acceptable TTO methods were when applied to SCRQoL states, as this is not the type of trade-off that might more realistically be considered with respect to health states. Because of the potential sensitivity of the issue, it was decided to limit this aspect of the study to members of the general population.

## Comparison of best–worst scaling and discrete choice experiment valuations

In order to determine the extent to which valuations of SCRQoL states obtained through a BWS were comparable to those obtained through DCE, we undertook a pilot study with 300 members of the general population. Nine domains were in the measure when we were designing the pilot study: *food and drink, personal cleanliness, accommodation, safety, social participation, occupation, control, dignity* and *living in own home* (later dropped – see *Chapter 2*). Each domain had four levels, except for *living in own home*, which had two levels (living in own home, not living in own home). For purposes of clarity, and in order to avoid wording that may lead to some domains dominating the choices, the *dignity* domain was described to respondents as ‘the way I am helped’ and the *employment and occupation* domain was presented as ‘use of my time’.

### Discrete choice experiments

The task of trading-off across nine different QoL domains in a single DCE would be particularly challenging. We drew on previous work that had used the approach of splitting domains across two DCEs (DCE1 and DCE2), with overlap in some domains to allow models from the two experiments to be based on a common utility scale. This study had demonstrated that the approach produced consistent values, and the utility parameters of overlapping domains were equal.<sup>10</sup>

**TABLE 9** Assessment of preference elicitation methods in the context of social care

Characteristic of method	VA	SG	TTO	DCE	BWS
Task complexity	Low	High	Medium	Medium	Low
Theoretically sound	No	Yes	Yes	Yes	Yes
Anchoring	Yes	Yes	Yes	Yes or no <sup>a</sup>	No
Assumption about preferences	Cardinal	Cardinal	Cardinal	Ordinal <sup>b</sup>	Ordinal <sup>b</sup>

<sup>a</sup> If a length of life attribute is included then anchoring to ‘being dead’ can be conducted; otherwise no.

<sup>b</sup> Although the responses from individuals are assumed to be ordinal, the estimated parameters have cardinal properties, as they represent functions of choice frequencies (on a probability scale).

The principles of grouping the domains into the two experiments were the following:

- *Control over daily life* and *dignity* should appear in both experiments.
- *Safety, personal care* and *food and drink* should appear in the same experiment, as these could be seen as core or basic outcomes of social care services.
- *Employment and occupation* and *social participation and involvement* should also appear in the same experiment.

The logic of this allocation of domains between the two experiments was, first, that people may make trade-offs between *personal care* and *safety*. An example from earlier cognitive interviews was a lady who would prefer to be clean and put herself at some risk getting into the bath, rather than be safe and unclean. The other reason is because *safety, personal care* and *food and drink* can be represented as the core outcomes of social services. Cleaning the house, social participation and being active/occupied can be seen as at a less fundamental level in terms of such scales as Maslov's hierarchy of needs.<sup>37</sup> By splitting the attributes across two experiments, we would not be able to explore any interactions in preferences for domains that were in different experiments, so it was necessary to make a priori assumptions about which interactions were likely to be most important when grouping the domains. The final allocation of the domains between the two choice experiments is shown in *Table 10*.

Both DCE1 and DCE2 used a forced D-efficient choice design and each of the design matrices included 128 situations. These matrices covered sufficient domain-level combinations to allow all two-way interactions between domain levels in the same experiment to be estimated. Obviously, 128 situations for each experiment would be too difficult for one respondent to evaluate. Therefore, the design matrices were divided into 16 orthogonal blocks, so that each respondent was presented with eight situations per choice experiment. Each choice scenario presents two situations as a list of domains at different levels. Examples of the introductory scripts and choice exercises in DCE1 and DCE2 are shown in *Box 3* and *Figure 5*, and *Box 4* and *Figure 6*, respectively.

### Best-worst scaling experiment

The BWS experiment used exactly the same attributes as the DCE but, rather than splitting them into two groups, all nine domains were presented in a single situation. Previous work had found that as many as nine domains in each situation is feasible as respondents are asked only to decide which of the domain levels presented is best and which is worst.<sup>97</sup> Choice situations in the BWS experiment were specified using an orthogonal main effects plan (OMEP). The full plan consisted of  $48 \times 21$  combinations, which were reduced to 32 situations in eight orthogonal blocks of four scenarios each. Each respondent was assigned a random sample of three blocks, so was asked to indicate his/her best and worst domain levels in 12 different situations. *Figure 7* shows an example of a BWS exercise, where the respondent was asked to choose the best and then the worst domain.

**TABLE 10** Grouping of domains between the two choice experiments

DCE1		DCE2	
1	<i>Food and drink</i>	4	<i>Accommodation cleanliness and comfort</i>
2	<i>Personal care</i>	5	<i>Social participation and involvement</i>
3	<i>Safety</i>	6	<i>Employment and occupation (use of my time)</i>
7	<i>Control over daily life</i>	7	<i>Control over daily life</i>
8	<i>The way I am helped (dignity)</i>	8	<i>Dignity (the way I am helped)</i>
		9	<i>Living in own home</i>

**BOX 3** Script introducing DCE1<sup>a</sup>

From this point on I would like you to imagine that you are in a situation where, through either illness or accident, you are no longer able to do everything you might expect to do for yourself without some assistance

I will ask you to look at a series of scenarios – in each we will present two different situations to you. These situations may have some good and bad aspects

In the following choices we will be concerned with these aspects of quality of life:

- food and drink
- personal care
- safety
- control over daily life
- the way I am helped

I would like you to weigh up the pros and cons of finding yourself in each situation. Then I would like you to tell me which situation you would consider to be the worst

These situations are imaginary, but I would like you to think about how you would feel if you were in these situations. There are no right or wrong answers to these choices; we are only interested in your views

<sup>a</sup> Scripts for DCE1 and DCE2 varied according to the order of appearance, e.g. DCE1 first and DCE2 second. Boxes 3 and 4 assume that DCE1 and DCE2 are presented first during the survey.

**Which of these two situations would you consider to be the worst?**

	Situation A	Situation B
<b>Food and drink</b>	<b>I can't always get all the food and drink I need, and I think there is a risk to my health</b>	<b>I can't always get all the food and drink I need, but I don't think there is a risk to my health</b>
<b>Personal care</b>	<b>I do not feel adequately clean or presentable</b>	<b>I feel adequately clean and presentable</b>
<b>Safety</b>	<b>Generally I feel as safe as I want</b>	<b>Most of the time I don't feel safe enough</b>
<b>Control over daily life</b>	I have as much control over my daily life as I want	I have as much control over my daily life as I want
<b>The way I am helped</b>	<b>The way I'm helped undermines the way I think and feel about myself</b>	<b>The way I'm helped sometimes undermines the way I think and feel about myself</b>
<b>Choice (mark 'X' in worst option)</b>	<input type="checkbox"/>	<input type="checkbox"/>

**FIGURE 5** A choice exercise example from DCE1.

In selecting the situations, it was important to consider the potential for bias resulting from differences in the size of the random utility component variance by attribute.<sup>102</sup> For example, presenting a situation in which one attribute was at its 'top' level (while all other attributes were at an intermediate level) was likely to make the 'best' choice easy (unless the levels of that attribute were all of moderate size and similar on the latent scale). Therefore, the random utility variance, ceteris paribus, would be small compared with a situation with all attributes at intermediate levels. As the OMEP was a small fraction of the full factorial, it was possible to avoid such problematic situations; the coding of domain levels was chosen to avoid designing situations defined by every attribute at its 'top' level, those with every attribute at its 'bottom' level and 'easy to choose' situations of the type described above.

**BOX 4** Script introducing DCE2

From this point on I would like you to imagine that you are in a situation where, through either illness or accident, you are no longer able to do everything you might expect to do for yourself without some assistance

I will ask you to look at a series of scenarios – in each we will present two different situations to you. These situations may have some good and bad aspects

In the following choices we will be concerned with these aspects of quality of life:

- accommodation cleanliness and comfort
- social participation and involvement
- use of my time
- control over daily life
- the way I am helped
- living in own home

I would like you to weigh up the pros and cons of finding yourself in each situation. Then I would like you to tell me which situation you would consider to be the worst

These situations are imaginary, but I would like you to think about how you would feel if you were in these situations. There are no right or wrong answers to these choices; we are only interested in your views

**Which of these two situations would you consider to be the worst?**

	Situation A	Situation B
<b>Accommodation cleanliness and comfort</b>	<b>My home is adequately clean and comfortable</b>	<b>My home is as clean and comfortable as I want</b>
<b>Social participation and involvement</b>	<b>I don't feel lonely and I have enough contact with people I like</b>	<b>I have as much contact as I want with people I like</b>
<b>Use of my time</b>	<b>I spend my time as I want, doing things I value or enjoy</b>	<b>I have enough things I value or enjoy to do with my time</b>
<b>Control over daily life</b>	I have some control over my daily life but not enough	I have some control over my daily life but not enough
<b>The way I am helped</b>	The way I'm helped does not make me think or feel any differently about myself	The way I'm helped does not make me think or feel any differently about myself
<b>Living in own home</b>	<b>And I am living in my own home</b>	And I am not living in my own home
<b>Choice (mark 'X' in worst option)</b>	<input type="checkbox"/>	<input type="checkbox"/>

**FIGURE 6** A choice exercise example from DCE2.

All respondents participated in the DCE1, DCE2 and BWS experiments. The order of the experiments was randomised in order to be able to test whether or not the order of appearance of experiments and, hence, whether or not fatigue and boredom or self-learning effects (i.e. the respondent gains more experience as he/she completes more exercises) had any effect on the responses of participants. There were four ordering possibilities, shown in *Table 11*, which were randomly allocated across respondents. This allowed development of models that enable testing differences, for example between responses where BWS exercises were presented first and BWS exercises were presented last.

Which of these nine points would you rate as being the best and which as being the worst?

Best	Aspect of life	Worst
	I can't always get all the food and drink I need, but I don't think there is a risk to my health	
	I feel adequately clean and presentable	
	My home is as clean and comfortable as I want	
	Sometimes I don't feel safe enough	
	I have as much contact as I want with people I like	
	I don't have enough things I value or enjoy to do with my time	
	I have some control over my daily life but not enough	
	The way I'm helped makes me think and feel better about myself	
	And I am not living in my own home	

FIGURE 7 Best-worst scaling choice exercise.

TABLE 11 Order patterns of appearance of the experiments

Pattern	Order of appearance		
	First	Second	Third
1	DCE1	DCE2	BWS
2	DCE2	DCE1	BWS
3	BWS	DCE1	DCE2
4	BWS	DCE2	DCE1

### Data collection and sample characteristics

The data collection involved house-to-house recruitment with the questionnaire administered through CAPI. The pilot was conducted in March 2009 among 300 adults in the south-east of England and Birmingham. The majority of the interviews were completed within approximately 30 minutes. However, some interviews, mostly among older respondents, took between 45 minutes and 1 hour.

The sample was not specified to be nationally representative, but rather the aim was to oversample ethnic minorities and those aged > 65 years to allow a more thorough review of the ability of these population segments to undertake the various choice tasks.

Details about this sample of 300 respondents who completed the interview are shown in *Table 12*. Older people (aged  $\geq 65$  years) and black and minority ethnic (BME) groups were sufficiently represented in the sample. In order to get a sense of the range of income levels, respondents were asked whether they received any benefits or tax credits. Out of the 300 participants, 29.3% indicated that they received benefits or tax credits, including income support (6%), income-based jobseeker's allowance (3%), working family tax credit (8%), disabled person's tax credit (4.6%),

**TABLE 12** Sample for methodological piloting

Sample	<i>n</i> =300
Aged ≥ 65 years (%)	48.3
Female (%):	48.7
■ Aged < 65 years	52.9
■ Aged ≥ 65 years	44.1
Ethnic non-white	26.3
Married	56.7
Working full-time (%):	22.7
■ Aged < 65 years	40.0
■ Aged ≥ 65 years	4.1
Retired (%):	50.3
■ Aged < 65 years	9.0
■ Aged ≥ 65 years	82.9

housing benefit (10%), council tax benefit (11%) or pension credits (10%). (Numbers may include multiple responses to different benefits or tax credits.)

In the DCEs, the alternatives were unlabelled, so we would expect to observe a balance in preferences between situations A and B. In DCE1, situation A was selected 1306 times (48%) and situation B was selected 1394 times (52%); in DCE2, situation A was chosen 1139 times (47.5%) and situation B was chosen 1260 times (52.5%). Preferences between situations A and B were not statistically different using the chi-squared test for equality of proportions [ $\chi^2 = 0.405$ ,  $p = 0.524$ , degrees of freedom (df) = 1]. The observed patterns implied that there were no unobserved biases towards a systematic preference for the left or right alternative.

Self-rating questions and interviewers' observations were used to develop criteria for excluding data from further analysis. In particular, respondents should have been able to:

- put themselves in an imaginary position
- understand the descriptions in the choices
- consider all aspects of choices
- feel that they were able to respond to the choices.

In addition, interviewers should have indicated that the respondent:

- understood the tasks 'a little', 'a great deal' or 'completely'
- gave the questions 'some', 'careful' or 'very careful' consideration
- did not lose concentration in the later stages of the interview.

Finally, the data set used in modelling excluded respondents who consistently chose situations A or B in all eight exercises in DCE1 or DCE2. *Table 13* shows the number of observations available for model development after excluding responses that failed to meet the above-mentioned criteria. Tests were undertaken on the models from each of the DCEs to examine whether or not the point estimates of the coefficients for each of the attribute levels changed significantly as a result of these exclusions. These tests suggested that any change in the point estimates was not significant at the 95% level, and the models with the exclusions applied were found to better fit the data (as measured through the rho-squared statistic).



**TABLE 13** Number of respondents excluded from the discrete choice analysis

Question	DCE1	DCE2	BWS
Could not put themselves into an imaginary position	42	42	42
Could not understand the descriptions in the choices	2	2	2
Did not look at all aspects of choices	1	1	1
Felt that they were unable to respond to the choices	2	2	2
Did not understand very much or at all	3	3	3
Gave the questions little or no consideration	5	5	5
Lost concentration in the later stages	2	2	2
Respondents that consistently chose the same alternative	3	5	0
<b>Total number of observations excluded<sup>a</sup></b>	<b>60</b>	<b>62</b>	<b>57</b>
<b>Total number of observations available for modelling</b>	<b>240</b>	<b>238</b>	<b>243</b>

a Numbers for each criterion do not add up to compute the total number of observations excluded, as more than one condition may apply to each observation.

### Testing for ordering effects

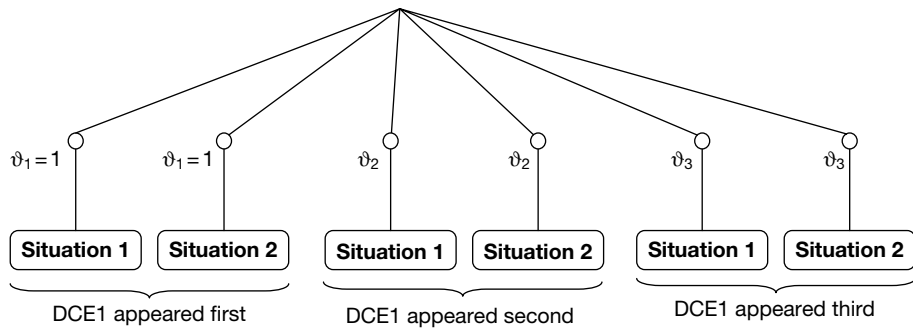
Prior to the development of the final models, we tested whether or not the order of appearance of the DCE1, DCE2 and BWS experiments (see *Table 11*) resulted in significant differences in the estimated models. We divided the sample in each model into three segments, depending on whether the experiment (i.e. DCE1, DCE2 or BWS) was completed first, second or third. For each segment we specified different scale parameters to reflect this ordering. In the case of BWS, only two scale parameters were specified reflecting whether the BWS was presented first or last.

As shown in *Figure 8*, the scale parameters of the first branch (i.e. DCE1 appears first) were fixed at the value of '1' for identification. The objective is to estimate the scale parameters  $\vartheta_2$  and  $\vartheta_3$ , after imposing the restriction that domain-level coefficients be equal across branches. If  $\vartheta_2$  and  $\vartheta_3$  are not significantly different from 1 then there is no evidence of a difference in the variance of the unexplained (unobserved) component of utility according to the experiment order. The same test can be applied for DCE2. As shown in *Figure 9*, the case of the BWS experiment requires only one scale parameter to be estimated, because the BWS appeared either first or last. *Table 14* shows the values of the scale parameters in the DCE1, DCE2 and the BWS data, along with their corresponding standard error (SE) and *t*-statistic values. The results indicate that scale coefficients are statistically significant from 0, but are not significantly different from 1. As a result, the order of appearance does not affect the level of noise within the responses.

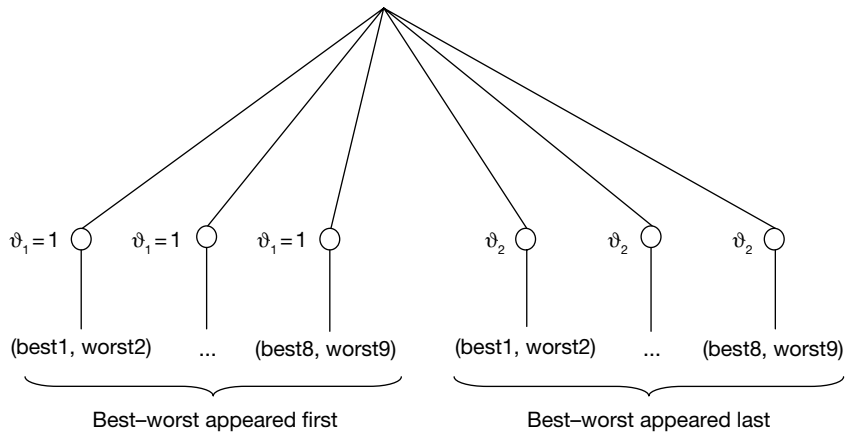
### Comparison of values between discrete choice experiment and best-worst scaling

The results from DCE1, DCE2 and the BWS experiments can be compared if we look at the marginal values of moving from the lowest level (e.g. *Dignity\_1*, which is constrained to 0) to the highest level (e.g. *Dignity\_2* and *Dignity\_4* in *Figure 10*) in each domain.

The models have different scales, hence the coefficients cannot be directly compared,<sup>105</sup> but we can look at the relative size of the differences by using one of the domain levels as a common denominator and scaling all others relative to this. In this case, we have chosen the highest level of need of the *control* domain (i.e. *Control\_4*), which was strongly estimated in all models. (All negative domain level coefficients have been divided by the negative coefficient of control, and that is why the coefficients of each domain level in *Figures 10–12* appear to be positive.) *Figures 10–12* provide a comparison of the relative values and SEs (see Hess and Daly<sup>106</sup> for computation of SEs). *Figure 10* presents the comparison of domain coefficients that were common across



**FIGURE 8** Testing differences in coefficient estimates because of the order of appearance of DCE1. Similar is the case for DCE2.



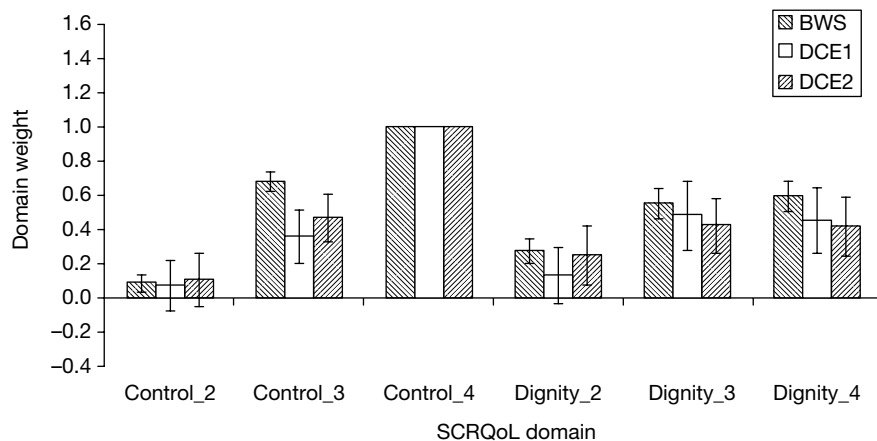
**FIGURE 9** Testing differences in coefficient estimates because of the order of appearance of BWS experiment.

**TABLE 14** Scale coefficients for testing the influence of experiments' order of completion

	$\vartheta_2$				$\vartheta_3$			
	Value	SE	t-statistic	Wald test <sup>a</sup>	Value	SE	t-statistic	Wald test
DCE1	1.14	0.11	10.0	1.27	1.11	0.15	7.6	0.73
DCE2	0.873	0.11	7.6	-0.12	1.22	0.17	7.1	1.29
BWS	1.04	0.04	28.9	1	-	-	-	-

<sup>a</sup> Wald test = value of parameter  $(\vartheta - 1)/SE$ . If the computed absolute value is smaller than the critical value  $\pm 1.96$ , then we cannot reject the hypothesis that the scaling parameters are statistically different to 1.

More detailed description of the econometric analysis and discussion of modelling results can be found in Potoglou *et al.*<sup>104</sup>



**FIGURE 10** Rescaled common domain weights across BWS, DCE1 and DCE2.

DCE1, DCE2 and BWS. Figure 11 shows the comparison of domain coefficients between BWS and DCE1, and Figure 12 shows the coefficients that appeared in DCE2 against those estimated in BWS.

The results revealed a broadly similar pattern in preferences using BWS and DCE. In 8 out of 11 cases, the domain-level weights common between DCE1 and BWS were not statistically different at the 95% significance level. Only the third level of *control* and second and third level of *food and drink* were statistically different at the 95% and 99.9% confidence intervals (CIs), respectively. In DCE2 and BWS, 8 out of 12 weights were not statistically different from each other at the 95% significance level. The rescaled BWS estimates were statistically different to those from the DCE2 for *employment and occupation* (level 3), *social participation* (level 3) and *living in own home*.

Finally, the values placed on *dignity* and *control* between the two separate DCE exercises (for which these were the common attributes) were not statistically different. As a result, the two different groupings in the design appear to support consistent valuations of these common attributes across respondents.

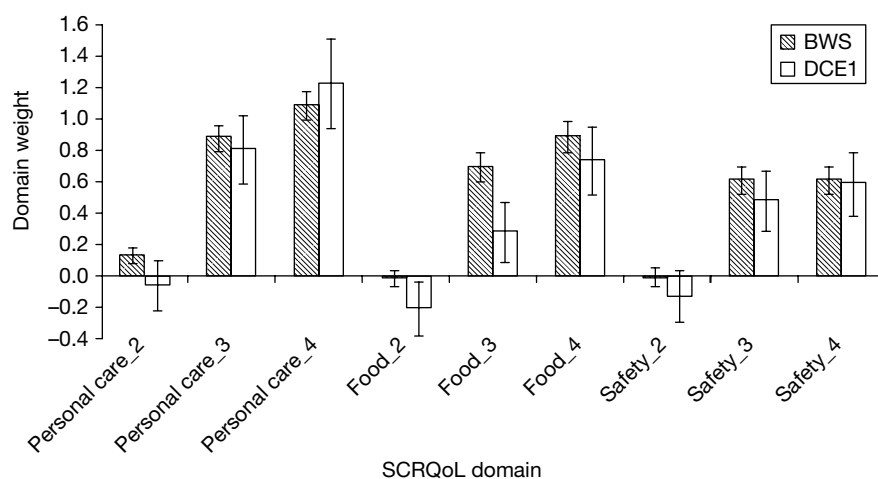


FIGURE 11 Rescaled common domain weights across BWS and DCE1.

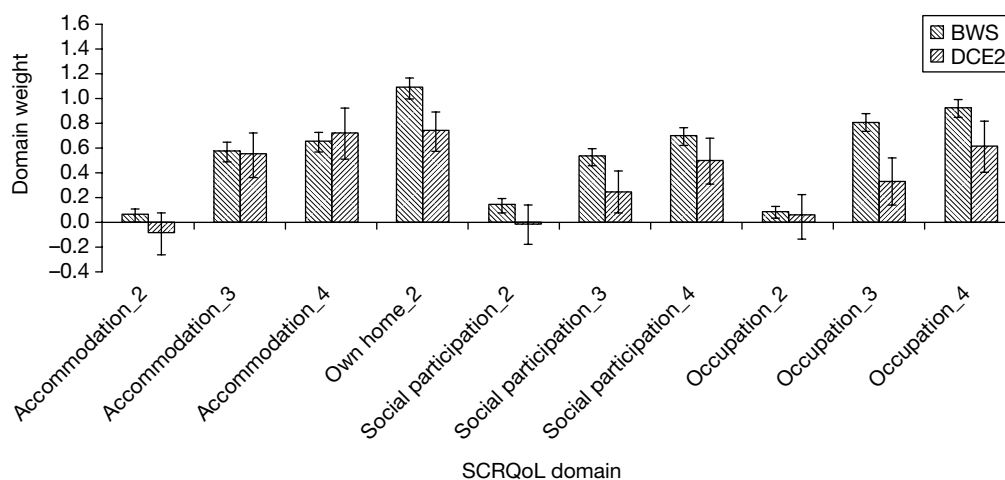


FIGURE 12 Rescaled common domain weights across BWS and DCE2.

## Service user cognitive testing

The results of the pilot preference study suggested that while there might be some advantages for BWS over DCE, in that all of the attributes or domains could be considered together, the resulting weights would not be dependent on the technique used. However, if we are to have confidence in the estimates, we need to be clear about the basis on which people are responding. In particular, we wanted to explore the acceptability and cognitive burden of this unusual type of questioning among service users. We wanted to find out how best to present the tasks to people, whether or not people understood what was expected of them, and whether or not they could manage the tasks.

The testing of the approaches to preference elicitation was conducted at the same time as the cognitive testing of the basic measure using the same methods (see *Chapter 2*). We presented users with the tasks as they would be presented in the actual study, and asked them to complete them as they would if the interview was part of a real study designed to elicit preferences. When giving their choices, we asked the interviewees to explain and give reasons for the choices they made.

### *Introductory wording and explaining the tasks*

We tried a number of explanatory wordings for both of the tasks before we found ones that seemed to work well. Short introductions seemed to work best. For the DCE task, we found it helpful to describe the choices in terms of the situations of imaginary people and to follow this up by asking interviewees which person they would prefer to be. This seemed to work better than asking interviewees ‘which of the imaginary situations A or B they would prefer to be in’ as it encouraged them to think about option A (and option B) as a single experience. We also found that formatting the options in the DCE in a way that encouraged people to scan down – rather than across – the page helped with understanding.

### *Understanding of the tasks, feasibility and acceptability*

Once the introductory wording and presentation of the tasks were simplified and clarified, the majority of service users found both the DCE and BWS tasks acceptable and feasible. That said, we did interview some people who found the imaginary nature of the task difficult. Some of the problems seemed to be related to their experience of needing care and having to respond to the failings of the service. A typical comment was ‘if someone treated me badly, I’d tell the person to go away so I wouldn’t be in that situation’. Another similar type of comment was ‘I would never have nothing to do, because I’d make sure I found something to do’. The interviewee who made the last comment then went on to say that having nothing to do would be terrible and really depressing, and this was effectively the rationale for avoiding the situation. When people reacted in this way to the descriptions, it affected the choices they made, and could lead to inaccuracies in the preferences. Interviewees who felt the situation could be avoided would often then exclude that situation from the choice set, and in the case of BWS choose another domain as the worst.

In one case, the interviewee lacked the cognitive capacity to complete the task (this person was able, however, to complete the basic measure asking about their current SCRQoL) and the interview had to be terminated. In a few instances, we did encounter interviewees who had difficulty with the questions, and one who refused to make a choice, arguing that it was not possible to imagine what life would be like. This person drew on his experience of his own illness and how this had affected his life, saying ‘since having my illness I’m a different person to who I was before’. Rather than missing out the questions, people who found it difficult to imagine a change to their life circumstances often chose the option in the DCE task that corresponded most to their current state. In the BWS exercise, people who found the tasks difficult generally

struggled to give any answer without a lot of support to help them navigate their way through the options. For example, the interviewers helped the interviewees to organise the statements into negative and positive statements, which made the task easier to manage. Interestingly, people who found it difficult to imagine themselves in the situations presented responded positively to a direct question and did not try to cover up their difficulties.

### **Heuristics and strategies used to make choices**

A benefit of cognitively testing the questions was that we gathered a large number of data on people's rationales for their preferences, and we were able from this to identify strategies that were commonly used. Some of these strategies were heuristic devices that are likely to produce inaccuracies in the preferences observed. Because of the differences in the tasks, the strategies differed between the DCE and the BWS tasks. We describe the variety of strategies used in both tasks.

In the DCE, as well as the strategy of choosing the option most similar to the interviewee's own SCRQoL taken by some people who struggled with the tasks, as we noted above, people also used other heuristic devices to make the task easier. One approach was to choose the option that was best for the first domain of SCRQoL presented in the situations. This approach would lead to biased preferences.

Another approach seen in the DCE was interviewees counting the number of negative and positive statements associated with each option and choosing the option that had the fewest negatives. This would produce biased estimates of preferences, as the assumed function form of the utility function is based on summing estimated attribute level part-worths. However, it is notable, when compared with the way the majority of people approached the task, as most people tended to focus on the wording of the domains and their meaning. For example, interviewees often preferred the option which had the best *control over daily life* state, even if the other domains associated with that choice were extremely negative, implying that preferences for *control* are close to being dominant in this context.

Interestingly, a common argument given for choosing the option which had the best *control over daily life* state was that, if they had control over their daily life then they could improve all of the other aspects. One interpretation of this strategy is that the interviewees find the option presented logically inconsistent. One interviewee stated this explicitly, saying 'If I had control over my life, then I wouldn't be in the position of X'. For these people, *control over daily life* seems to be logically prior to the other domains, and causes some problems when weighing up options. This difficulty is unique to the DCE task, as it requires the domains to be presented in a logically coherent situation.

In the BWS task, the length of the list seemed to make the task quite difficult. To make the task easier, some people seemed to focus on the options towards the top of the list and ignore those towards the end of the list. As mentioned above, people who found the imaginary nature of the task difficult ignored those domains that troubled them. Both of these strategies would lead to inconsistent preferences. However, the former would only be identified this way if the ordering of the domains was changed between the questions. Interestingly, people who tended to have these difficulties also tended to answer the questions much more quickly than those who understood the task fully.

As we saw in the DCE task, in the BWS task some domains dominated. For example, it was common for people to choose *control over daily life* as the best option if it was at any of the levels except high-level needs (when it was chosen as the worst option). Interviewees argued that, without control, none of the other domains was attainable: as one interviewee explained,

‘everything flows from this [control over daily life]’. Another domain that dominated was *living in own home*, which was dropped from the measure. As explained in *Chapter 2*, we dropped this domain because when people explained their choices and what they understood by living in their own home (or not living in their own home), they described their situations solely in terms of the other SCRQoL domains. Furthermore, as discussed elsewhere, the inclusion of an attribute that dominates all others requires a different conceptual framework for estimating preferences, as the model assumptions are violated.<sup>103</sup>

Although many people chose the *control over daily life* domain consistently as the best or worst, depending on the state represented, this was not true of everyone. We noticed that best–worst choices often reflected the personality and experiences of the interviewee. For example, one man with mental health problems was very nervous socially, and felt lonely and socially isolated. He tended to select the *social participation* and *safety* options as worst or best, as they were particularly salient to his condition and life more generally.

As we describe below, this process proved very helpful in considering how best to design the main stage for eliciting preferences for both service users and the general population, and helped to inform the analysis.

## Time trade-off

The basic BWS and DCE models generate values on the latent scale for all social care states. The problem is how to anchor these on to the ‘QALY’ scale where ‘0’ represents ‘being dead’ and ‘1’ represents the social care equivalent of ‘perfect health’. One solution to this problem is to value the worst state described by a classification using a cardinal preference elicitation technique such as TTO, and use this to ‘anchor’ the BWS or DCE values on the QALY scale. This relies on using just one data point and has been shown to poorly predict mean TTO values for other health states.<sup>107</sup> A better approach would be to map from BWS to TTO using more data points and estimating the functional form.

The objective is to identify the point at which QoL states are viewed by respondents as equal to being dead – and to scale the utility scores accordingly. However, we needed to know whether or not such an approach would be acceptable and understandable for members of the general public in the context of social care rather than health-care outcomes. Moreover, we wanted to test using a computer-based format, rather than with a physical board and cards, which has been the predominant method used in the past.

The TTO exercise was designed following established protocols.<sup>93,94</sup> Respondents were asked to choose between having *all* of their social care needs met for a period of < 10 years, followed by immediate death (Life A); and having a lower QoL for 10 years (Life B). The point of trade-off is established by varying the duration of the ‘all needs met’ state (Life A), and asking respondents to indicate at which point they view Life A and Life B to be equal. An example of a scenario is shown in *Figure 13*.

Four waves of interviews were conducted. The first two waves asked 10 respondents to consider six SCRQoL states. These interviews helped to identify a number of useful changes, including:

- amendments to the introduction to the interview, to improve the presentation of the TTO tasks
- an increase from six to eight ‘evaluation states’ for the third and fourth waves of the testing, given that participants did not appear to be distressed, fatigue or bored when evaluating six states, and were likely to be able to evaluate more states

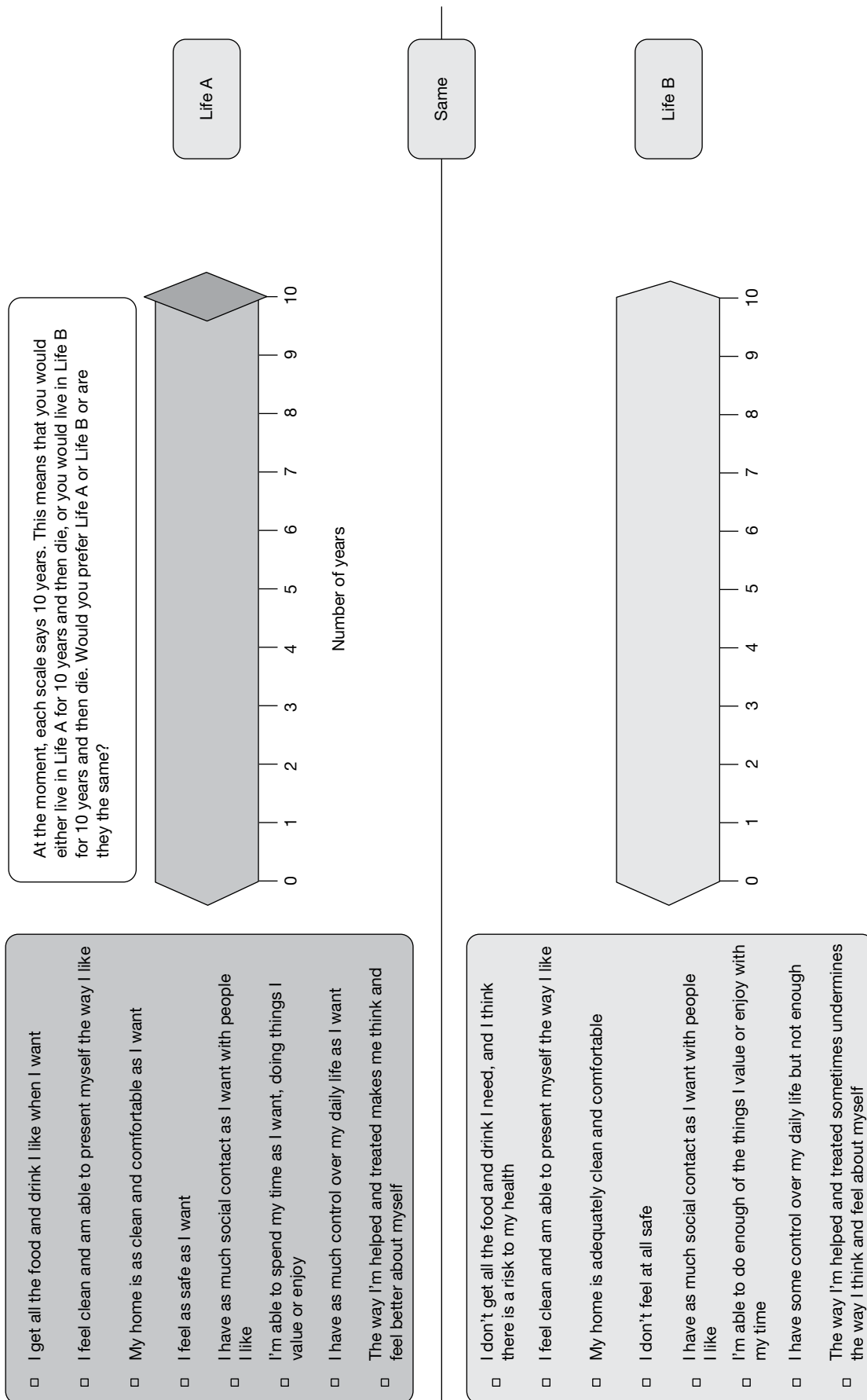


FIGURE 13 An example valuation scenario in the TTO experiment. Note: The upper state was subsequently replaced by 'all needs met'.

- minor revisions to the screen layout and choice instructions to assist in the clarity of the task
- a change in the format of Life A (best state), from a list of all domains at the ‘no needs’ level to a simple statement of ‘all needs met’, supported by a prior explanation of the levels associated with this ‘all needs met’ state in the introduction.

In total, 19 respondents were interviewed across the four waves of the cognitive testing. Slightly over half of the sample (11) was female, most (16) were white, and a range of age groups and socioeconomic groups were represented.

Data collected during the interviews were used to review general trends in the behaviour of individuals in these TTO choice tasks to examine whether their behaviour aligned with expectations and to identify issues which might require additional consideration. Exploratory analysis examined the relationship between the TTO results and the utility scores based on BWS. [These were based on the results of the MOPSU preference study<sup>10</sup> (see *Table 1*).] In principle, we would expect that the TTO and utility scores would follow a monotonic relationship, i.e. the TTO score would increase as the calculated BWS utility score increased. In the majority of the cases, we were able to observe just such a monotonic relationship, although there were a number of outliers.

In the first two waves, three respondents valued mid-score utility states as being worse than being dead, suggesting that they either had extreme preferences or, more likely, had misunderstood the TTO tasks. This issue may have been resolved via the developing understanding of the interviewer, as only one respondent in the third and fourth waves responded in this way.

It was also notable that, in the third and fourth waves, three respondents had a uniform (and high) distribution of TTO scores, regardless of the state under evaluation. It may be that participants were unwilling to trade-off any time from their life – regardless of state – with being dead. Feedback from the cognitive interviews suggested that this was the case for at least one respondent, who refused to sacrifice any time in Life A to avoid Life B, and consistently chose Life B for 10 years.

## Implications for the main stage design

To our knowledge, the pilot exercise we undertook was the first to test empirically the comparability of the profile-case BWS and DCE estimates. The results were encouraging. Overall, we observed similar patterns between the DCE and BWS, and in the majority of cases the estimated preference weights (when normalised/rescaled) were not statistically different. The differences that were observed in the BWS and DCE estimates might reflect bias in the latter caused by different respondents inferring different information about the attributes they did not ‘see’: omitting attributes from a DCE typically affects estimates of both the attribute level means and the variance (scale).<sup>108</sup> In BWS, respondents saw all of the attributes at once. The tests for preference equality suggested such effects at the aggregate level may be small. However, there may be important differences among subgroups. Any such differences would have implications for the generalisability of the results to the wider population. An important lesson for the main stage, therefore, was that the analysis should investigate – and, if necessary, control for – heterogeneity across different segments of the population (e.g. education level, place of residence).

In comparing the DCE and BWS approaches, BWS has the advantage that respondents are presented with all domains at the same time. BWS also provides more information than a traditional DCE for a given set of choices, as coefficients are estimated on a common scale (allowing the measurement of the relative value of the domains and not just differences within



each domain). Moreover, BWS appeared to pose a smaller cognitive burden than a DCE in which respondents need to weigh up the combined value of all the attributes simultaneously. This was consistent with the findings in the wider literature (e.g. see Flynn *et al.*<sup>102</sup>).

We felt it was important to investigate, and ideally reflect, the service user perspective in weighting our measure. Although the cognitive testing with service users identified some potential problems in establishing preferences, once the introduction and layout were clear we found that, overall, people were able to undertake the tasks. We established an approach for identifying people for whom the task was too cognitively demanding, and strategies for identifying and investigating the incidence and impact of strategies, such as taking into account only certain domains.

The results of the TTO testing were also encouraging. Respondents appeared to find the questions understandable and acceptable and, after some modifications, the computer-aided approach worked well. The lack of respondent fatigue suggested that we could increase the number of states presented from the minimum, which was desirable as there were potential non-linearities in respondents' preferences across the range.<sup>68</sup> In addition, the exercise identified the importance of interviewer briefing and training for this type of interview, which helped inform the main stage processes.

As we described in *Chapter 1*, we took advantage of parallel studies to feed into this project. The preference study, conducted as part of the MOPSU research, is described in more detail elsewhere.<sup>109</sup> Key aspects from the perspective of the current study were that it allowed us to test the BWS approach, gathering further information about 'second'-best and worst attributes in each scenario (see *Chapter 5*), with a revised, close-to-final version of the ASCOT measure, which excluded the *living in own home* domain with a general population sample. In addition to allowing us to extend the sample size for the main BWS analysis, this piece of work allowed us to generate weighted ASCOT states for the main stage TTO evaluation.



## Chapter 5

# Development of utility weights

### Introduction

Having finalised the design of the ASCOT instrument (see *Chapter 3* and *Appendix 5*) and being informed by the results of the feasibility work on preference elicitation (see *Chapter 4*), we turn now to the main preference elicitation surveys and the development of utility weights for our measure. We start by describing the design of the BWS experiments, and the surveys and data on which we were able to draw for the main analysis. We present the analysis undertaken on the BWS data and the resulting domain weights for both the general population and the service user samples. We then turn to the design and results of the TTO experiments. In *Chapter 6* we then provide a step-by-step example of how the utility weighting scheme can be applied by analysts to calculate a comparative score for different SCRQoL states, and ultimately a corresponding social care equivalent QALY score.

### Design of the best–worst scaling experiment

The final version of the ASCOT measure has eight domains, each with four levels. As in the pilot study, the BWS tasks were developed using a fractional–factorial design, which reduced the full factorial plan (four<sup>8</sup> tasks) to a matrix of 32 tasks. This design matrix was blocked using a procedure that sought to minimise the correlation of the levels being presented for the domains within the block (i.e. avoiding all respondents in one version of the survey always seeing a given domain at the same level). An additional consideration in defining the BWS tasks was the minimisation of ‘easy options’ or straightforward choices in each task. To achieve this, the domain levels were re-coded in the design matrices through an iterative procedure to ensure that each BWS task had more than one domain at the end point levels (i.e. level 1 or 4), thereby generating situations where non-obvious choices between domains were required. Each respondent was given eight BWS tasks.

*Figure 14* shows an example of a BWS exercise used. In addition to choosing the best and worst domains, respondents were also asked to specify their second-best and second-worst choices. Specifically, the respondents indicated the best domain from the list of eight domains. The chosen domain was excluded from the list, and respondents were asked to choose the worst domain from the shortened list. The chosen domain was again excluded from the list and, following the same pattern, respondents were asked to indicate the second-best and second-worst domains of those remaining. Asking for second-best and second-worst choices reduces the effects of lexicographic and non-trading behaviour. Further, it is possible to gain additional information on the importance of the other issues in the absence of a domain that was first chosen as best or worst.<sup>89,110</sup> Finally, it is possible to draw more information per choice task as we obtain a partially complete ranking of four out of eight domains.

The order of the attributes was varied for each respondent in order to prevent ordering bias and to be able to separate the effect of domain choice from the position of that domain within a task.

Aspect of life	
1	My home is as clean and comfortable as I want
3	I don't always get adequate or timely food and drink
4	I don't feel at all clean or presentable
5	I have some control over my daily life but not enough
6	I have some social contact with people but not enough
7	The way I'm helped and treated sometimes undermines the way I think and feel about myself
8	I don't do anything I value or enjoy with my time



Aspect of life	
1	My home is as clean and comfortable as I want
2	I feel as safe as I want
3	I don't always get adequate or timely food and drink
4	I don't feel at all clean or presentable
5	I have some control over my daily life but not enough
6	I have some social contact with people but not enough
7	The way I'm helped and treated sometimes undermines the way I think and feel about myself
8	I don't do anything I value or enjoy with my time

(a) Which of these eight aspects would you rate as being the best?



(b) Which of the remaining seven aspects would you rate as being the worst?

Aspect of life	
1	My home is as clean and comfortable as I want
3	I don't always get adequate or timely food and drink
5	I have some control over my daily life but not enough
6	I have some social contact with people but not enough
7	The way I'm helped and treated sometimes undermines the way I think and feel about myself
8	I don't do anything I value or enjoy with my time



(d) Which of the remaining five aspects would you rate as being the next worst?

(c) Which of the remaining six aspects would you rate as being the next best?

FIGURE 14 Choice sequence in the BWS task.

This BWS design was tested in the MOPSU study with members of the general population and it worked well. In that study, we used the version of the measure that was tested with older home care service users, which, as described in *Chapter 3*, was modified slightly as a result of the findings from both studies:

- *accommodation cleanliness and comfort* at level 3, and
- *food and drink* at levels 2, 3 and 4.

Although the MOPSU and OSCA surveys were conducted some time apart (in 2009 and 2010, respectively), we sought to maintain as much consistency as possible in their design so we could pool the data to provide a larger general population sample on which to base the final BWS-based utility weights.

### The samples

The OSCA general population survey was undertaken in August 2010, with 500 respondents located in Birmingham, Newcastle, London and the south-east of England. The MOPSU study was conducted in October and November 2009 in the same areas and had 1000 respondents, about half of whom (513) had used the four-level version of ASCOT. The key characteristics of respondents of both surveys compared with the general population, as reported in the 2001 census, are provided in *Appendix 6*. This shows that the samples in both the OSCA and MOPSU surveys were very similar, and broadly consistent with the general population in all of the observable dimensions.

The OSCA service user sample was recruited in August to November 2010 through nine councils, drawing on respondents to the 2010 UES, which was conducted of people who had received equipment services. The final sample included 456 people, who were broadly representative of respondents to the UES nationally. Key demographic and socioeconomic characteristics are shown in *Appendix 6*.

### Respondents' ability to undertake best-worst scaling tasks

The survey included a number of control questions in order to determine whether or not respondents were able to answer the best-worst experiments, and specifically whether or not they:

- understood the descriptions in the choices
- felt that they were able to respond to the choices.

Interviewers also provided an assessment of the respondents' ability to answer the questions, by indicating whether or not respondents:

- understood what he/she was asked to do in the questions
- put thought into responding
- maintained concentration during the survey.

*Table 15* shows the number of respondents who provided negative responses in the control questions, along with interviewers' observations about respondents' ability to get involved with the survey.

The criteria set for including an observation in the discrete choice analysis were:

- respondents should have been able to:
  - understand the descriptions in the choices
  - feel able to respond to the choices, and

**TABLE 15** Summary of responses in the control questions

Control questions	OSCA				MOPSU	
	General population		Service users		No. of respondents	%
	No. of respondents	%	No. of respondents	%		
Could not understand the descriptions in the choices	9	1.8	53	11.6	17	3.3
Felt that they were unable to respond to the choices	5	1	51	11.1	16	3.1
1. Did not understand at all	3	0.6	9	2.0	10	1.9
2. Did not understand very much	4	0.8	26	5.7	13	2.5
3. Understood a little	45	9.0	131	28.6	63	12.3
4. Understood a great deal	111	22.2	140	30.6	151	29.4
5. Understood completely	336	67.2	152	33.2	275	53.6
No response was provided	1	0.2	0	0	1	0.2
1. Gave the questions no consideration	1	0.2	7	1.5	4	0.8
2. Gave the questions little consideration	8	1.6	40	8.7	28	5.5
3. Gave the questions some consideration	53	10.6	95	20.7	88	17.2
4. Gave the questions careful consideration	126	25.2	148	32.3	155	30.2
5. Gave the questions very careful consideration	311	62.2	168	36.7	237	46.2
No response was provided	1	0.2	0	0	1	0.2
1. Lost concentration in the later stages	1	0.2	10	2.2	9	1.8
2. Lessened concentration in the later stages	19	3.8	48	10.5	37	7.2
3. Maintained concentration with a deal of effort throughout the survey	34	6.8	93	20.3	71	13.8
4. Maintained concentration with some effort throughout the survey	97	19.4	129	28.2	130	25.3
5. Easily maintained concentration throughout the survey	348	69.6	178	38.9	265	51.7
No response was provided	1	0.2	0	0	1	0.2

- interviewers should have indicated that the respondent:
  - understood the tasks at least a little
  - gave the questions some, careful or very careful consideration
  - did not lose concentration in the later stages of the survey.

Finally, we excluded respondents who consistently chose the first (or second, third, etc.) of the list of attributes within a task as being best across all eight exercises offered to them.

*Table 16* summarises the total number of observations that were excluded from the modelling when these rules were applied sequentially. Not surprisingly given the cognitively demanding nature of the task, a higher proportion of service users were excluded than the general population.

### **Modelling the best–worst choice data**

The analyses of the BWS general population and service user data were conducted using discrete choice models, refining initial models, and testing hypotheses and associations before settling on final models to use as an estimate for our preference weights:

**TABLE 16** Observations to be excluded from the BWS modelling

Control questions	OSCA		
	General population	Service users	MOPSU
Could not understand the descriptions in the choices	9		17
Felt unable to respond to the choices	3	51	12
Did not understand very much or at all	5	18	16
Gave the questions little or no consideration	3	21	12
Lost concentration in the later stages	0	1	2
Non-traders	1		2
Incomplete data	1		1
Total number of observations excluded <sup>a</sup>	22 (5%)	91 (25%)	62 (14%)
Total number of observations available for modelling	478	367	451

a Numbers for each criterion do not add up to compute the total number of observations excluded as more than one condition may apply to each observation.

- Estimation of SCRQoL domain weights after controlling for the order of domains in a given BWS task, sample level variance heterogeneity for best/worst responses, and excessive randomisation (in terms of the presentation of the tasks as a result of a fieldwork problem – described further in the following section) in the MOPSU data set.
- Estimation of domain weights after controlling for the order of domains in a given BWS task, variance heterogeneity for best/worst responses, excessive randomisation in the MOPSU data set, and with the additional consideration of variance heterogeneity occurring across different segments in the sample.
- Comparison of general population and service user domain weights based on the second set of models.
- Estimation of domain weights after controlling for the order of domains in a given BWS task, variance heterogeneity for best/worst responses, excessive randomisation in the MOPSU data set, variance heterogeneity occurring across different segments in the sample, and with the additional consideration of socioeconomic characteristics of respondents.
- Estimation of domain weights when the four-level domains used in this study collapsed into three levels by combining levels 1 and 2 in each domain and after controlling for the order of domains in a given BWS task, variance heterogeneity for best/worst responses, excessive randomisation in the MOPSU data set, and variance heterogeneity occurring across different segments in the sample (reported in *Appendix 7*).

*Table 17* shows the number of observations available for modelling in the general population and the service users' data sets.

Initially we developed multinomial logit models (MNLs)<sup>58,90</sup> under the assumption that the best, worst, second-best and second-worst choices made by the same respondents were independent of each other (although from a diminishing set of choices). Also, the initial models were estimated without explicit consideration of the correlation between observations from the same individual. This model structure was considerably simpler to set up and had very short run times; thus, allowing a large number of different model specifications to be tested before moving to a more complex models. While this was a simplification with a potential for correlation within the responses from a given individual, both within a given BWS task and between BWS tasks, these correlations were accounted for in the final models, which are reported in the following sections.

**TABLE 17** Number of observations available for modelling in the general population and service users samples

Data	No. of choices per best-worst task (best, worst, second best, second worst)	No. of best-worst tasks per respondent	No. of respondents	No. of observations available for modelling
General population (OSCA and MOPSU)	4	8	929	29,728
Service users	4	8	367	11,744

Each domain that could have been chosen within the best-worst task was specified as a separate alternative within the choice model, and was given a utility function that took account of the level at which that domain was presented to the respondent in the task in question. The structure of this model is shown in *Figure 15*.

The MNL model was specified such that the utility function for a given domain was defined as:

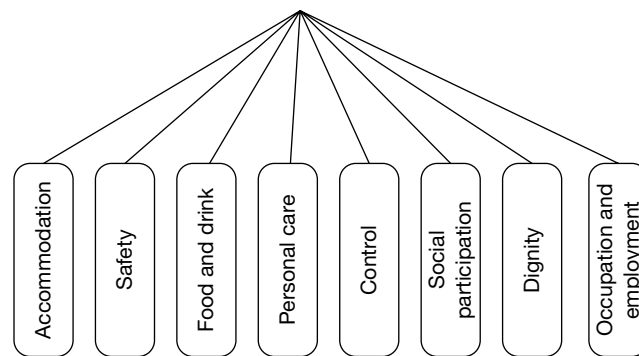
- a linear-additive function of the products between the weights (coefficients) to be estimated and the dummy coding of the corresponding domain levels (only one of which will ever take a value of 1 for any given choice), and
- a number of dummy-coded variables to control for the position of a domain level in the best-worst task when that domain level was chosen as being best, worst, second best or second worst.

Some initial tests were undertaken to assess whether or not it was feasible to pool the data from the MOPSU and the main OSCA general population surveys, thereby providing the greatest return from the investment in data collection. Separate models were estimated from the data sets from the two waves, which both returned well estimated and highly significant coefficients. However, given that these were independently drawn samples, which were given similar but slightly different BWS designs, we had to allow for the potential for differences in scale between the two models. Therefore, rather than directly comparing the coefficients from the two models we looked at the ratio with respect to Control level 1 (which had a consistent definition in each instrument and was the most strongly estimated parameter in both models).

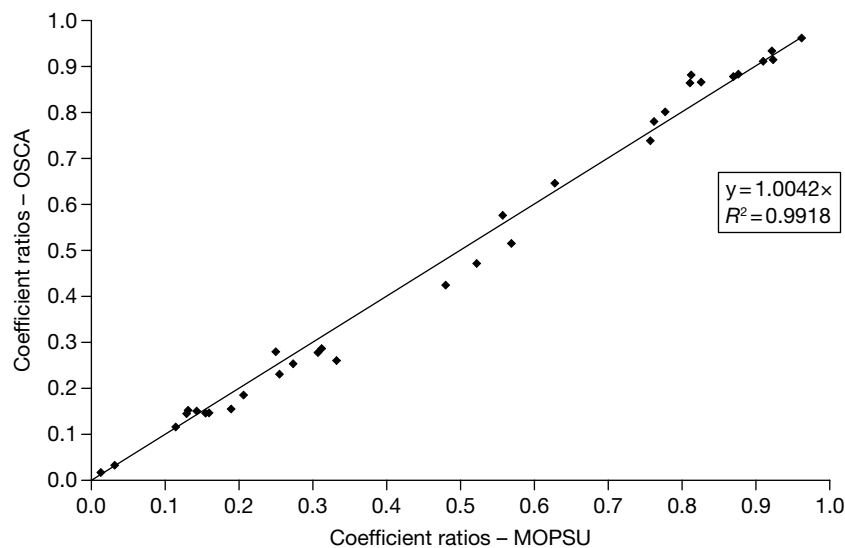
*Figure 16* illustrates the extent of the consistency between the results obtained from the independent modelling of the two data sets, and shows that an  $R^2$  of 0.992 is achieved when the scaled coefficient ratios from the two models are regressed against each other. Having established that the preferences were highly consistent (once the differences in model scale were taken into account) we decided to pool the data from the two survey waves. This necessitated the move to a scale heterogeneity MNL (S-MNL) model structure (using the naming convention developed by Fiebig *et al.*<sup>93</sup>), which allows difference in model scale between data sets or individuals. In this case, the model was specified to allow heterogeneity in scale between the two data sets, allowing for the possibility that one of the data sets might contain less error than the other; however, in subsequent models this structure was also used to account for the difference in model scale between different groups of respondents.

In order to take advantage of the pooled data, it was necessary to estimate separate coefficients for the different definitions used for *accommodation cleanliness and comfort* and *food and drink* in the MOPSU and OSCA data sets. The coefficients corresponding to the other domain levels could be jointly estimated from the two data sets. Most importantly, preliminary estimations showed that the model scale (amount of noise in the choice data) was not significantly different between the two data sets. As a result, this level of freedom was dropped from the model specification, which halved the number of utility functions and reduced the model run time.





**FIGURE 15** Model structure in the analysis of the BWS data.



**FIGURE 16** Comparison of relative parameter estimates from MOPSU and OSCA.

Owing to an error in the MOPSU preference survey (see *Table 1*) development, and miscommunication between the market research agency and their interviewers, 190 people or 37% of the sample in the MOPSU survey were given a version of the BWS experiments where the order of the domains was also randomised within the BWS tasks for an individual. This is referred to as 'excessive randomisation'. There was more confusion (noise) among respondents who were given the excessive-randomisation BWS tasks, but the point estimates of domain-level weights were not, on the whole, significantly different, once the scale difference between the data sets was taken into account.<sup>109</sup> As the impact of the excessive randomisation was found to be relatively small, it was possible to specify the models with an additional scale parameter to allow for differences in the consistency, which these groups may have demonstrated in their choices within the exercises (with the a priori assumption from the MOPSU findings that those with greater randomisation in the task would have been less consistent in the reporting of their preferences).

Preliminary analysis with the pooled data set also found strong evidence of a difference in scale between 'best' and 'worst' choices, with people appearing to find it easier to identify the best, rather than the worst, aspect of the situations. As a result, the models reported below also control for potential differences in respondents' ability to choose the 'best' and 'worst' domains. In addition to common problems, our prior expectation was that some respondents might find it easier to respond to BWS tasks than others. These differences between groups in the sample were likely to introduce more noise (higher variance) in the models, which needed to be captured

using appropriate model specifications. This was confirmed through tests that indicated that there were differences associated with socioeconomic characteristics and place of residence. A detailed discussion of the rationale and final models' specification is provided in *Appendix 8*.

Finally, the presentation of the BWS tasks in both the general population and the service user surveys involved randomisation of the order of domains between respondents in order to control for the impact that the position of the domain might have had on the likelihood of that domain being selected as the best or worst option. This randomisation ensured that, if the position of the domain was to induce any bias, all of the domains would suffer this bias equally and there would be no systematic over- or underestimation of the value of any given domain. By recording the position of each domain in any given BWS task, it was possible to incorporate this information within the parameterisation of the model and quantify the increase (or decrease) in the likelihood of the respondents choosing a given domain owing to its position independent of the estimation of the value placed on the level at which was presented. The final models contained eight position dummy-coded variables, which reflected the position of a domain level when that was chosen, contingent upon whether the respondent was asked to choose a 'best' or 'worst' option.

A generalised example of the utility function specification for the *accommodation cleanliness and comfort* domains is shown in *Box 5*.

## Best-worst scaling results

### Overall preferences

#### General population

*Table 18* presents the domain-level weights estimated for the general population using the combined OSCA and MOPSU data. The goodness-of-fit measure as indicated by the rho-squared value = 0.229 showed that the model performed relatively well.<sup>111</sup> (A rho-squared value between 0.25 and 0.30 of a discrete choice model is roughly equivalent to a value of between 0.75 and 0.80 of a linear regression model, which is considered a model with a good fit.) All domain levels were estimated relative to domain *control* at level 4, which was defined as the base level because it had the lowest utility: respondents felt that having no *control over their daily life* was the worst domain level of all those presented. All other domain levels are valued more highly so have positive weights compared with this. In the tables the model coefficients that relate to the wording used in the MOPSU instrument are shown in italic text.

The estimated weights in *Figure 17* were all statistically significant, with relatively high *t*-ratios.

An examination of the coefficient estimates and their 95% CIs in *Figure 17* revealed a small number of cases where the levels within the same domain were not statistically different from each other:

- *food and drink* between levels 1 and 2 in the MOPSU data set
- *occupation* between levels 1 and 2, and
- *dignity* between levels 3 and 4.

These findings imply that respondents either could not identify the differences in the wording, or felt that the outcomes of these domain levels were the same. However, it was encouraging that the *food and drink* weights at levels 1 and 2 in the OSCA experiment were significantly different, suggesting that the problem identified earlier had been successfully addressed by changes in the wording.

**BOX 5** Generalised form of the utility function for the domain *accommodation cleanliness and comfort*

$$\begin{aligned}
U_i(\text{accommodation}) = & \\
& + \beta_{\text{accomOSCA}} \times (1, \text{ if accommodation level}^a = 1)_i \times (1, \text{ if choice scenario} = \text{best or second best})_i \times (1, \text{ if OSCA data}) \\
& - \beta_{\text{accomOSCA}} \times (1, \text{ if accommodation level} = 1)_i \times (1, \text{ if choice scenario} = \text{worst or second worst})_i \times (1, \text{ if OSCA data}) \\
& + \beta_{\text{accomOSCA}} \times (1, \text{ if accommodation level} = 2)_i \times (1, \text{ if choice scenario} = \text{best or second best})_i \times (1, \text{ if OSCA data}) \\
& - \beta_{\text{accomOSCA}} \times (1, \text{ if accommodation level} = 2)_i \times (1, \text{ if choice scenario} = \text{worst or second worst})_i \times (1, \text{ if OSCA data}) \\
& + \beta_{\text{accomOSCA}} \times (1, \text{ if accommodation level} = 3)_i \times (1, \text{ if choice scenario} = \text{best or second best})_i \times (1, \text{ if OSCA data}) \\
& - \beta_{\text{accomOSCA}} \times (1, \text{ if accommodation level} = 3)_i \times (1, \text{ if choice scenario} = \text{worst or second worst})_i \times (1, \text{ if OSCA data}) \\
& + \beta_{\text{accomOSCA}} \times (1, \text{ if accommodation level} = 4)_i \times (1, \text{ if choice scenario} = \text{best or second best})_i \times (1, \text{ if OSCA data}) \\
& - \beta_{\text{accomOSCA}} \times (1, \text{ if accommodation level} = 4)_i \times (1, \text{ if choice scenario} = \text{worst or second worst})_i \times (1, \text{ if OSCA data}) \\
& + \beta_{\text{accomMOPSU}} \times (1, \text{ if accommodation level} = 1)_i \times (1, \text{ if choice scenario} = \text{best or second best})_i \times (1, \text{ if MOPSU data}) \\
& - \beta_{\text{accomMOPSU}} \times (1, \text{ if accommodation level} = 1)_i \times (1, \text{ if choice scenario} = \text{worst or second worst})_i \times (1, \text{ if MOPSU data}) \\
& + \beta_{\text{accomMOPSU}} \times (1, \text{ if accommodation level} = 2)_i \times (1, \text{ if choice scenario} = \text{best or second best})_i \times (1, \text{ if MOPSU data}) \\
& - \beta_{\text{accomMOPSU}} \times (1, \text{ if accommodation level} = 2)_i \times (1, \text{ if choice scenario} = \text{worst or second worst})_i \times (1, \text{ if MOPSU data}) \\
& + \beta_{\text{accomMOPSU}} \times (1, \text{ if accommodation level} = 3)_i \times (1, \text{ if choice scenario} = \text{best or second best})_i \times (1, \text{ if MOPSU data}) \\
& - \beta_{\text{accomMOPSU}} \times (1, \text{ if accommodation level} = 3)_i \times (1, \text{ if choice scenario} = \text{worst or second worst})_i \times (1, \text{ if MOPSU data}) \\
& + \beta_{\text{accomMOPSU}} \times (1, \text{ if accommodation level} = 4)_i \times (1, \text{ if choice scenario} = \text{best or second best})_i \times (1, \text{ if MOPSU data}) \\
& - \beta_{\text{accomMOPSU}} \times (1, \text{ if accommodation level} = 4)_i \times (1, \text{ if choice scenario} = \text{worst or second worst})_i \times (1, \text{ if MOPSU data}) \\
& + \beta_{\text{top}} \times (1, \text{ if accommodation appeared in the first row})_i \times (1, \text{ if choice scenario} = \text{best or second best})_i \\
& + \beta_{\text{second top}} \times (1, \text{ if accommodation appeared in the second row})_i \times (1, \text{ if choice scenario} = \text{best or second best})_i \\
& + \beta_{\text{second bottom}} \times (1, \text{ if accommodation appeared in the eighth row})_i \times (1, \text{ if choice scenario} = \text{best or second best})_i \\
& + \beta_{\text{bottom}} \times (1, \text{ if accommodation appeared in the ninth row})_i \times (1, \text{ if choice scenario} = \text{best or second best})_i \\
& + \beta_{\text{top}} \times (1, \text{ if accommodation appeared in the first row})_i \times (1, \text{ if choice scenario} = \text{best or second worst})_i \\
& + \beta_{\text{second top}} \times (1, \text{ if accommodation appeared in the second row})_i \times (1, \text{ if choice scenario} = \text{best or second worst})_i \\
& + \beta_{\text{second bottom}} \times (1, \text{ if accommodation appeared in the eighth row})_i \times (1, \text{ if choice scenario} = \text{best or second worst})_i \\
& + \beta_{\text{bottom}} \times (1, \text{ if accommodation appeared in the ninth row})_i \times (1, \text{ if choice scenario} = \text{best or second worst})_i \\
& + \varepsilon_i
\end{aligned}$$

where  $\varepsilon_i$  is the error term that captures the unobserved heterogeneity owing to differences among observations

a The '0-otherwise' statement was specified in all 'if-statements' of the utility functions, but has been omitted in the above boxes to save space in the document.

It was also informative to review the rank order of the estimated weights, as this ranking revealed the relative importance of each of the domain levels. As shown in *Figure 17*, the highest weight-value was estimated for the domain *control over daily life* at its top level ('I have as much control over my daily life as I want'). Closely following *control*, respondents then preferred the top two levels of *occupation and employment* ('I'm able to spend my time as I want' and 'I'm able to do

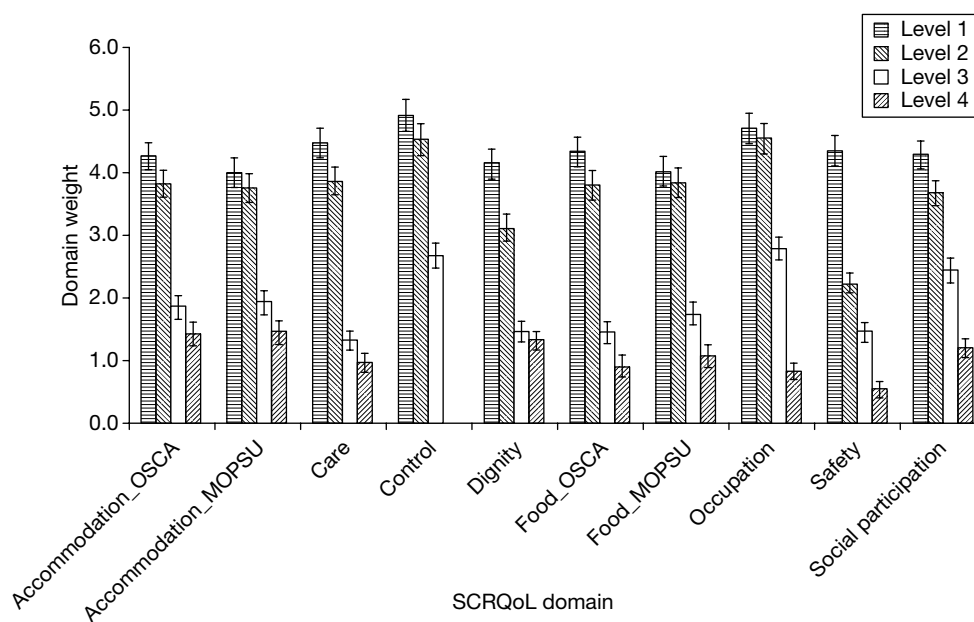
**TABLE 18** Estimated parameters using general population OSCA and MOPSU data

Domain level	Coefficient	t-ratio (robust)
<b><i>Accommodation cleanliness and comfort</i></b>		
1. My home is as clean and comfortable as I want	4.27	38.2
2. My home is adequately clean and comfortable	3.83	34.4
3. My home is not quite clean or comfortable enough	1.87	18.8
4. My home is not at all clean or comfortable	1.44	15.5
1. My home is as clean and comfortable as I want (MOPSU)	4.01	33.4
2. My home is adequately clean and comfortable (MOPSU)	3.77	33.0
3. My home is less than adequately clean or comfortable (MOPSU)	1.94	19.0
4. My home is not at all clean or comfortable (MOPSU)	1.47	16.5
<b><i>Safety</i></b>		
1. I feel as safe as I want	4.37	37.6
2. Generally I feel adequately safe, but not as safe as I would like	2.23	24.8
3. I feel less than adequately safe	1.47	18.6
4. I don't feel at all safe	0.558	7.9
<b><i>Food and drink</i></b>		
1. I get all the food and drink I like when I want	4.35	37.7
2. I get adequate food and drink at OK times	3.82	32.3
3. I don't always get adequate or timely food and drink	1.47	16.1
4. I don't always get adequate or timely food and drink, and I think there is a risk to my health	0.932	10.2
1. I get all the food and drink I like when I want (MOPSU)	4.04	33.0
2. I get food and drink adequate for my needs (MOPSU)	3.85	33.1
3. I don't get all the food and drink I need, but I don't think there is a risk to my health (MOPSU)	1.76	18.8
4. I don't get all the food and drink I need, and I think there is a risk to my health (MOPSU)	1.09	12.0
<b><i>Personal care</i></b>		
1. I feel clean and am able to present myself the way I like	4.49	37.6
2. I feel adequately clean and presentable	3.88	35.6
3. I feel less than adequately clean or presentable	1.34	17.0
4. I don't feel at all clean or presentable	0.982	13.4
<b><i>Control over daily life</i></b>		
1. I have as much control over my daily life as I want	4.93	37.9
2. I have adequate control over my daily life	4.54	37.0
3. I have some control over my daily life, but not enough	2.69	25.9
4. I have no control over my daily life	0.00	Base level
<b><i>Social participation and involvement</i></b>		
1. I have as much social contact as I want with people I like	4.31	36.8
2. I have adequate social contact with people	3.69	34.3
3. I have some social contact with people, but not enough	2.46	24.7
4. I have little social contact with people and feel socially isolated	1.22	16.2
<b><i>Dignity</i></b>		
1. The way I'm helped and treated makes me think and feel better about myself	4.16	35.0
2. The way I'm helped and treated does not affect the way I think or feel about myself	3.13	29.6
3. The way I'm helped and treated sometimes undermines the way I think and feel about myself	1.48	18.2
4. The way I'm helped and treated completely undermines the way I think and feel about myself	1.33	17.1

**TABLE 18** Estimated parameters using general population OSCA and MOPSU data (*continued*)

Domain level	Coefficient	t-ratio (robust)
<b>Occupation and employment</b>		
1. I'm able to spend my time as I want, doing things I value or enjoy	4.72	37.5
2. I'm able to do enough of the things I value or enjoy with my time	4.56	36.4
3. I do some of the things I value or enjoy with my time, but not enough	2.8	28.3
4. I don't do anything I value or enjoy with my time	0.843	12.5
<b>Domain position in the BWS task</b>		
Top: respondent chose the domain that appeared first	0.151	4.5
Top two: respondent chose the domain that appeared second	0.0755	2.7
Bottom: respondent chose the attribute that appeared last	0.0121	0.4
Bottom two: respondent chose the attribute that appeared before the last	-0.034	-1.2
<b>Scale parameters</b>		
$\mu_{\text{excessive}}$ : MOPSU data with 'excessive randomisation'	0.806	4.26 <sup>a</sup>
$\mu_{\text{worst}}$ : data relate to a 'worst' or 'second-worst' choice	0.843	7.4
<b>Model diagnostics</b>		
No. of observations	29,728	
df	45	
Final log-likelihood	-42,491.7	
Rho <sup>2</sup> (0)	0.229	

a For all scale parameters reported, the *t*-statistics reflect the hypothesis test of the scale being statistically different from 1. The ranking of the domain-level weights within each domain was as expected, with weights at the top levels (i.e. levels representing better SCRQoL) higher than weights in the second, third and fourth levels within each domain.

**FIGURE 17** Estimated weights and CIs in the general population model.

enough of the things I value or enjoy with my time'). This ranking of domain levels meant that, if respondents were presented with a BWS task in which all domains were at their top levels, they were more likely to choose the top level of *control over daily life*. When comparing the middle levels of domains, respondents in the general population again placed the highest valuation on the domain *control* ('I have adequate control over my daily life'), followed by *personal care* ('I feel adequately clean and presentable') and *accommodation cleanliness and comfort*. The least preferred levels were the lowest levels of *safety*, followed by *occupation and employment, food and drink* (OSCA) and *personal care*.

The parameters of the position variables capturing instances in which a domain was chosen when it appeared first or second from the top (i.e. top and top two) were also statistically significant. This finding implied that respondents were more likely to choose a domain (either as 'best' or 'worst') when it appeared first or second in the BWS task. The order implied by the coefficient values was intuitive; however, the overall effect of these variables as seen from the values of their coefficients was considerably smaller than the main effect of the domain weights, and the significance of these coefficients was also significantly lower. On the other hand, the parameters reflecting choices of domains when these appeared at the bottom or second from the bottom in a BWS task were not statistically significant.

Finally, the scale parameters  $\mu_{\text{worst}}$  and  $\mu_{\text{excessive}}$  were both significantly different from 1. The significance of the scale parameter  $\mu_{\text{worst}}$  implied that the variance (noise) between 'best' (including 'best' and 'second best') and 'worst' (including 'worst' and 'second worst') responses was significantly different. Having set the scale of 'best' choices equal to 1 for identification, the estimated value of scale equal to 0.843 implied higher variance in the 'worst' choices, which meant that the data from the 'worst' responses generated a less-deterministic choice process than the 'best' responses. Practically, this meant that respondents placed more emphasis on the domain levels when they indicated their 'best' rather than their 'worst' choices. The significance of the scale parameter  $\mu_{\text{excessive}}$  suggested that the data from those respondents facing excessive randomisation of BWS tasks had a less-deterministic choice process than those facing the intended randomisation of the domain order, which was held constant for a given respondent. Again, the value of the scale parameter being  $< 1$  implied that the 'excessive' randomisation data in the MOPSU data set contained more noise than the data with 'intended' randomisation.

### Service users

The domain-level weights estimated from the service users' BWS data are presented in *Table 19*. As in the general population models, the domain *control* at level 4 was set as the base level. All domain weights were statistically significant, and the model diagnostics showed that the service users' model also performed well.

*Figure 18* shows the ranking of domain-level weights within domains, which appeared to be consistent and agreed with prior expectations, with weights at top levels higher than weights in the second, third and fourth levels. Within domains, there were three cases where the levels within the same domain were not statistically different at the 95% level:

- *food and drink* between levels 1 and 2
- *occupation and employment* between levels 1 and 2
- *dignity* between levels 3 and 4.

The rank order of the estimated weights, as seen from *Figure 18*, revealed that *control over daily life* at level 1 ('I have as much control over my daily life') was also the aspect of SCRQoL that was valued the highest by service users. The second and third highest valued domains were *occupation and employment* and *personal care*, both at level 1. *Occupation and employment* and

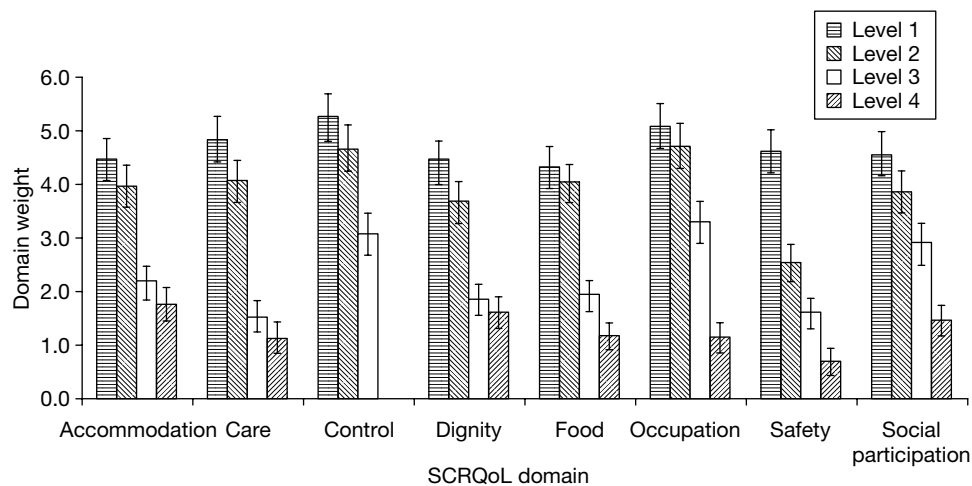
**TABLE 19** Estimated parameters using service user data

Domain level	Coefficient	t-ratio (robust)
<b><i>Accommodation cleanliness and comfort</i></b>		
1. My home is as clean and comfortable as I want	4.50	22.1
2. My home is adequately clean and comfortable	4.00	20.8
3. My home is not quite clean or comfortable enough	2.20	13.2
4. My home is not at all clean or comfortable	1.80	11.3
<b><i>Safety</i></b>		
1. I feel as safe as I want	4.66	22.5
2. Generally I feel adequately safe, but not as safe as I would like	2.57	15
3. I feel less than adequately safe	1.62	11.1
4. I don't feel at all safe	0.72	5.4
<b><i>Food and drink</i></b>		
1. I get all the food and drink I like when I want	4.34	21.4
2. I get adequate food and drink at OK times	4.04	21.4
3. I don't always get adequate or timely food and drink	1.96	12.9
4. I don't always get adequate or timely food and drink, and I think there is a risk to my health	1.19	9.3
<b><i>Personal care</i></b>		
1. I feel clean and am able to present myself the way I like	4.88	23.0
2. I feel adequately clean and presentable	4.09	20.3
3. I feel less than adequately clean or presentable	1.55	10.2
4. I don't feel at all clean or presentable	1.16	8.2
<b><i>Control over daily life</i></b>		
1. I have as much control over my daily life as I want	5.29	23.2
2. I have adequate control over my daily life	4.71	21.4
3. I have some control over my daily life, but not enough	3.09	15.2
4. I have no control over my daily life	0.00	Base level
<b><i>Social participation and involvement</i></b>		
1. I have as much social contact as I want with people I like	4.59	21.9
2. I have adequate social contact with people	3.90	19.1
3. I have some social contact with people, but not enough	2.92	14.4
4. I have little social contact with people and feel socially isolated	1.49	10.1
<b><i>Dignity</i></b>		
1. The way I'm helped and treated makes me think and feel better about myself	4.45	21.2
2. The way I'm helped and treated does not affect the way I think or feel about myself	3.70	18.5
3. The way I'm helped and treated sometimes undermines the way I think and feel about myself	1.88	12.5
4. The way I'm helped and treated completely undermines the way I think and feel about myself	1.63	10.7
<b><i>Occupation and employment</i></b>		
1. I'm able to spend my time as I want, doing things I value or enjoy	5.12	24.2
2. I'm able do enough of the things I value or enjoy with my time	4.74	21.9
3. I do some of the things I value or enjoy with my time, but not enough	3.32	16.4
4. I don't do anything I value or enjoy with my time	1.17	8.2

*continued*

**TABLE 19** Estimated parameters using service user data (*continued*)

Domain level	Coefficient	t-ratio (robust)
<i>Domain position in the BWS task</i>		
Top: respondent chose the domain that appeared first	0.272	5.2
Top two: respondent chose the domain that appeared second	0.138	3.1
Bottom: respondent chose the attribute that appeared last	0.0385	0.8
Bottom two: respondent chose the attribute that appeared before the last	-0.0451	-1.0
<i>Scale parameter</i>		
$\mu_{\text{worst}}$ : data relate to a 'worst' or 'second-worst' choice	0.6862	10.0
Model diagnostics		
No. of observations	11,744	
df	36	
Final log-likelihood	-17,042.5	
Rho <sup>2</sup> (0)	0.217	

**FIGURE 18** Estimated weights and CIs in the service users' model.

*control over daily life* at level 2 were the highest-valued middle-level domains and, overall, among the top five valued domains. Service users valued least the lowest levels of *safety* ('I don't feel at all safe'), *personal care* ('I don't feel at all clean or presentable'), *occupation and employment* ('I don't do anything I value or enjoy with my time') and *food and drink* ('I don't always get adequate or timely food and drink', and 'I think there is a risk to my health').

The parameters of the position variables were significant only in the case that the chosen domain appeared first or second in the BWS task. Service users were also more likely to choose a domain that appeared first or second in the domain list. The order implied by the value of the coefficients of the top and top two parameters was intuitive; however, the main effect of the domain-level weights and the significance of these coefficients were significantly lower. These findings were in line with the findings in the general population models.

At this stage, the service users' model included a single scale to account for potential differences between 'best' and 'worst' choices, as no service user had faced excessive randomisation. The estimated scale parameter  $\mu_{\text{worst}}$  was statistically significantly different with respect to 1



( $t$ -ratio = 10, see *Table 20*). This finding implies that there was more noise (higher variance) in the 'worst' choices relative to 'best' choices.

### Other factors associated with variance heterogeneity

A second set of models was estimated in which we further explored the potential influence of other scaling effects, known as unobserved or variance heterogeneity, in both the general population and service user data. Variances may differ for a number of reasons. The most intuitive reason for the presence of variance heterogeneity is that different people may have different degrees of certainty in their answers in the BWS task. Failure to model differences in variances leads to bias in the point estimates (utility weights),<sup>93,94</sup> and Flynn *et al.*<sup>96</sup> discuss this with reference to QoL data. The risk is that variations in preferences can be attributed to groups, which are about differences in scale consistency (or certainty) rather than differences in actual preferences.

In undertaking the analysis, we made a number of hypotheses of what we might expect to affect variations in scale consistency. In particular, we anticipated that scale consistency could be influenced by or associated with:

- education
- age
- experience of services and providing care
- current level of QoL
- location
- interviewer
- length of time taken during the exercises
- a person's ability to relate to the situation.

### General population

A series of model specifications was tested prior to arriving at the final model specification, which is reported in *Table 20*. The initial set of analyses investigated each of the potential associations of scale consistency in the model. The statistically significant scale parameters were added to the final model. The specification of the model with respect to domain levels and position-related variables remained the same as in the model reported in *Table 18*.

The scale parameters reported in the model shown in *Table 20* suggested that, in addition to the previous groups identified, four further subgroups of respondents in the general population sample made less deterministic choices than their counterparts:

- respondents who could not put themselves in the imaginary situations described
- respondents who lived in London or the south-east of England
- respondents with fair, bad or very bad self-reported health status
- where the interview time was shorter than 15 minutes.

*Figure 19* shows that the ranking of domain levels within the domains was retained. As expected, the values of the domain-level weights increased relative to the model with fewer scale parameters. This was expected because the remaining unobserved variance in the model became smaller after introducing the additional scale parameters.

In this model, however, we did notice an additional overlap between levels 1 and 2 of the domain *accommodation cleanliness and comfort*, which were previously found to be marginally different from each other. The overlaps between *food and drink* (MOPSU), *occupation and employment* at levels 1 and 2, and *dignity* at levels 3 and 4 were retained.

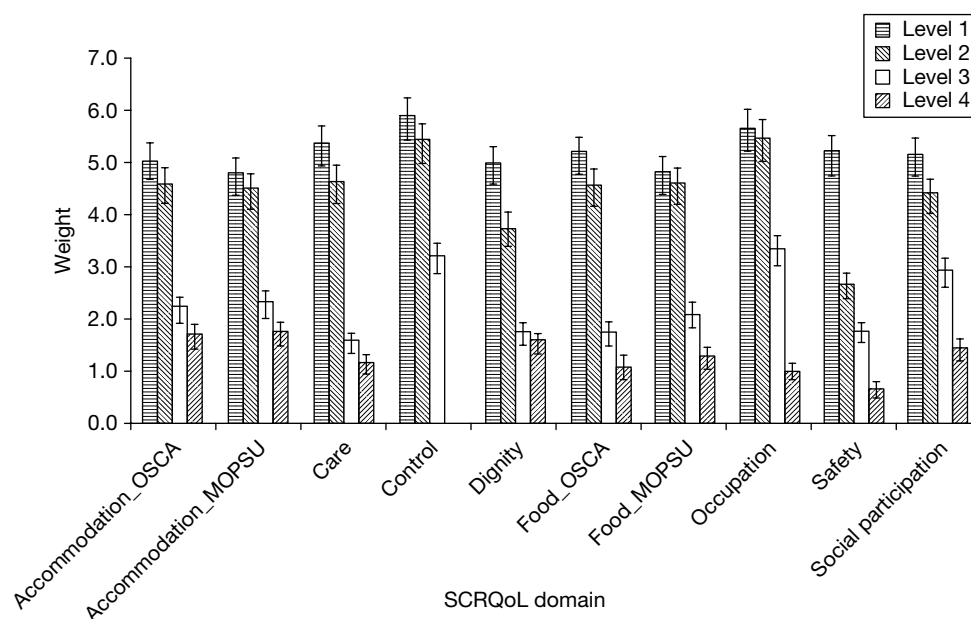
**TABLE 20** Estimated parameters using general population OSCA and MOPSU data with scale effects

Domain level	Coefficient	t-ratio (robust)
<b><i>Accommodation cleanliness and comfort</i></b>		
1. My home is as clean and comfortable as I want	5.06	27.3
2. My home is adequately clean and comfortable	4.57	25.9
3. My home is not quite clean or comfortable enough	2.19	16.8
4. My home is not at all clean or comfortable	1.68	14.1
1. My home is as clean and comfortable as I want (MOPSU)	4.76	25.9
2. My home is adequately clean and comfortable (MOPSU)	4.47	25.6
3. My home is less than adequately clean or comfortable (MOPSU)	2.28	17.1
4. My home is not at all clean or comfortable (MOPSU)	1.73	15.0
<b><i>Safety</i></b>		
1. I feel as safe as I want	5.16	27.1
2. Generally I feel adequately safe, but not as safe as I would like	2.65	20.3
3. I feel less than adequately safe	1.75	16.3
4. I don't feel at all safe	0.67	7.7
<b><i>Food and drink</i></b>		
1. I get all the food and drink I like when I want	5.15	27.4
2. I get adequate food and drink at OK times	4.54	25.1
3. I don't always get adequate or timely food and drink	1.72	14.5
4. I don't always get adequate or timely food and drink, and I think there is a risk to my health	1.08	9.8
1. I get all the food and drink I like when I want (MOPSU)	4.77	25.2
2. I get food and drink adequate for my needs (MOPSU)	4.55	24.9
3. I don't get all the food and drink I need, but I don't think there is a risk to my health (MOPSU)	2.09	16.6
4. I don't get all the food and drink I need, and I think there is a risk to my health (MOPSU)	1.26	11.4
<b><i>Personal care</i></b>		
1. I feel clean and am able to present myself the way I like	5.34	26.9
2. I feel adequately clean and presentable	4.62	25.8
3. I feel less than adequately clean or presentable	1.55	15.3
4. I don't feel at all clean or presentable	1.14	12.6
<b><i>Control over daily life</i></b>		
1. I have as much control over my daily life as I want	5.86	27.1
2. I have adequate control over my daily life	5.39	26.7
3. I have some control over my daily life, but not enough	3.17	21.3
4. I have no control over my daily life	0.00	Base level
<b><i>Social participation and involvement</i></b>		
1. I have as much social contact as I want with people I like	5.12	27.2
2. I have adequate social contact with people	4.38	25.8
3. I have some social contact with people, but not enough	2.91	20.1
4. I have little social contact with people and feel socially isolated	1.41	14.6
<b><i>Dignity</i></b>		
1. The way I'm helped and treated makes me think and feel better about myself	4.96	25.6
2. The way I'm helped and treated does not affect the way I think or feel about myself	3.73	22.7
3. The way I'm helped and treated sometimes undermines the way I think and feel about myself	1.73	16.6
4. The way I'm helped and treated completely undermines the way I think and feel about myself	1.54	15.7

continued

**TABLE 20** Estimated parameters using general population OSCA and MOPSU data with scale effects (*continued*)

Domain level	Coefficient	t-ratio (robust)
<b>Occupation and employment</b>		
1. I'm able to spend my time as I want, doing things I value or enjoy	5.64	26.5
2. I'm able do enough of the things I value or enjoy with my time	5.43	26.1
3. I do some of the things I value or enjoy with my time, but not enough	3.32	22.3
4. I don't do anything I value or enjoy with my time	1.00	11.9
<b>Domain position in the BWS task</b>		
Top: respondent chose the domain that appeared first	0.18	4.6
Top two: respondent chose the domain that appeared second	0.08	2.5
Bottom: respondent chose the attribute that appeared last	0.02	0.4
Bottom two: respondent chose the attribute that appeared before the last	-0.04	-1.2
<b>Scale parameters</b>		
$\mu_{\text{excessive}}$ : MOPSU data with 'excessive randomisation'	0.82	3.7
$\mu_{\text{worst}}$ : data relate to a 'worst' or 'second-worst' choice	0.85	7.0
$\mu_{\text{not-imagine}}$ : data relate to individuals who could not put themselves in the imaginary positions described in the choices	0.79	3.2
$\mu_{\text{London and South-East}}$ : data relate to individuals living in London and the South East	0.82	5.3
$\mu_{\text{bhealth}}$ : data relate to individuals with self-reported fair, bad or very bad health	0.87	3.3
$\mu_{\text{int. duration}<15\text{min}}$ : data relate to individuals' interview duration shorter than 15 minutes	0.87	2.6
<b>Model diagnostics</b>		
No. of observations	29,728	
df	49	
Final log-likelihood	-42,327.9	
Rho <sup>2</sup> (0)	0.233	

**FIGURE 19** Estimated weights and CIs in the general population model.

The overall rank order of the estimated domain weights between the scale effects and limited scale effects models was retained, and the significance of the position parameters when the chosen domain appeared first or second in the BWS task was retained. Again, the estimates were intuitive and implied that a domain was more likely to be chosen if it appeared first or second. However, the main effect and significance of these variables were less than that of the domain levels.

### Service users

The specification tests for potential scale effects in the service users' sample revealed two further subgroups of respondents who made fewer deterministic choices:

- residents outside the south-east of England, and
- respondents who had 'no formal qualification' or just 'GCSE/O-level'.

Figure 20 shows that the ranking of domain levels within the domains was retained. As in the general population scale-effect model, the value of each domain-level weight increased compared with the limited scale effects model.

In the scale effects model, the top two and bottom two levels of the domain *accommodation cleanliness and comfort* were similarly valued by the respondents and were not statistically different, respectively. As with the limited scale effects model, the top two levels of *food and drink* and *occupation and employment* and the bottom two levels of *dignity* were not significantly different in this scale effects model.

With a few minor variations, the overall ranking of the domain-level weights found in the limited scale effects model was broadly retained, and the position parameters when the chosen domain appeared first or second in the BWS task were statistically significant as in the limited scale effects models.

**TABLE 21** Estimated parameters using service users' data with scale effects

Domain level	Coefficient	t-ratio (robust)
<b><i>Accommodation cleanliness and comfort</i></b>		
1. My home is as clean and comfortable as I want	6.48	13.6
2. My home is adequately clean and comfortable	5.74	13.1
3. My home is not quite clean or comfortable enough	3.14	9.9
4. My home is not at all clean or comfortable	2.58	9.3
<b><i>Safety</i></b>		
1. I feel as safe as I want	6.79	13.0
2. Generally I feel adequately safe, but not as safe as I would like	3.60	10.9
3. I feel less than adequately safe	2.25	8.9
4. I don't feel at all safe	1.01	5.2
<b><i>Food and drink</i></b>		
1. I get all the food and drink I like when I want	6.25	13.0
2. I get adequate food and drink at OK times	5.81	13.0
3. I don't always get adequate or timely food and drink	2.78	10.4
4. I don't always get adequate or timely food and drink, and I think there is a risk to my health	1.59	7.8

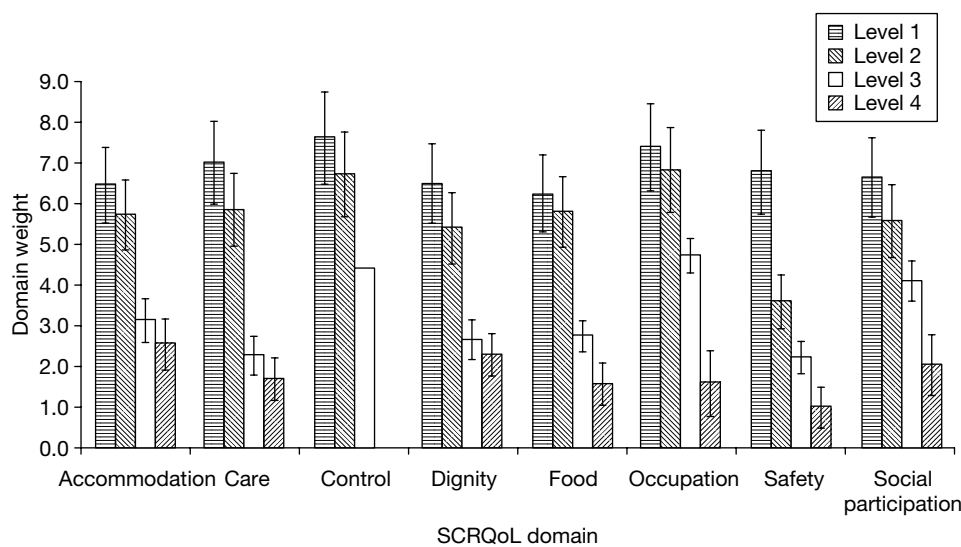
**TABLE 21** Estimated parameters using service users' data with scale effects (*continued*)

Domain level	Coefficient	t-ratio (robust)
<b>Personal care</b>		
1. I feel clean and am able to present myself the way I like	7.02	13.7
2. I feel adequately clean and presentable	5.87	12.8
3. I feel less than adequately clean or presentable	2.30	8.8
4. I don't feel at all clean or presentable	1.72	7.3
<b>Control over daily life</b>		
1. I have as much control over my daily life as I want	7.64	13.3
2. I have adequate control over my daily life	6.74	12.8
3. I have some control over my daily life, but not enough	4.42	11.2
4. I have no control over my daily life	0.00	Base level
<b>Social participation and involvement</b>		
1. I have as much social contact as I want with people I like	6.66	13.2
2. I have adequate social contact with people	5.59	12.2
3. I have some social contact with people, but not enough	4.11	10.8
4. I have little social contact with people and feel socially isolated	2.05	8.4
<b>Dignity</b>		
1. The way I'm helped and treated makes me think and feel better about myself	6.48	12.9
2. The way I'm helped and treated does not affect the way I think or feel about myself	5.41	12.0
3. The way I'm helped and treated sometimes undermines the way I think and feel about myself	2.68	10.2
4. The way I'm helped and treated completely undermines the way I think and feel about myself	2.30	9.4
<b>Occupation and employment</b>		
1. I'm able to spend my time as I want, doing things I value or enjoy	7.39	13.5
2. I'm able do enough of the things I value or enjoy with my time	6.84	13.0
3. I do some of the things I value or enjoy with my time, but not enough	4.75	11.4
4. I don't do anything I value or enjoy with my time	1.60	7.4
<b>Domain position in the BWS task</b>		
Top: respondent chose the domain that appeared first	0.37	4.9
Top two: respondent chose the domain that appeared second	0.17	2.6
Bottom: respondent chose the attribute that appeared last	0.04	0.6
Bottom two: respondent chose the attribute that appeared before the last	-0.06	-0.8
<b>Scale parameters</b>		
$\mu_{\text{worst}}$ : data relate to a 'worst' or 'second-worst' choice	0.69	9.5
$\mu_{\text{notSE}}$ : data relate to residents outside the south-east of England	0.66	7.9
$\mu_{\text{O-level}}$ : data relate to individuals with no formal qualification or GCSE/O-level	0.83	2.9
<b>Model diagnostics</b>		
No. of observations	11,744	
df	38	
Final log-likelihood	-16,914.7	
Rho <sup>2</sup> (0)	0.224	

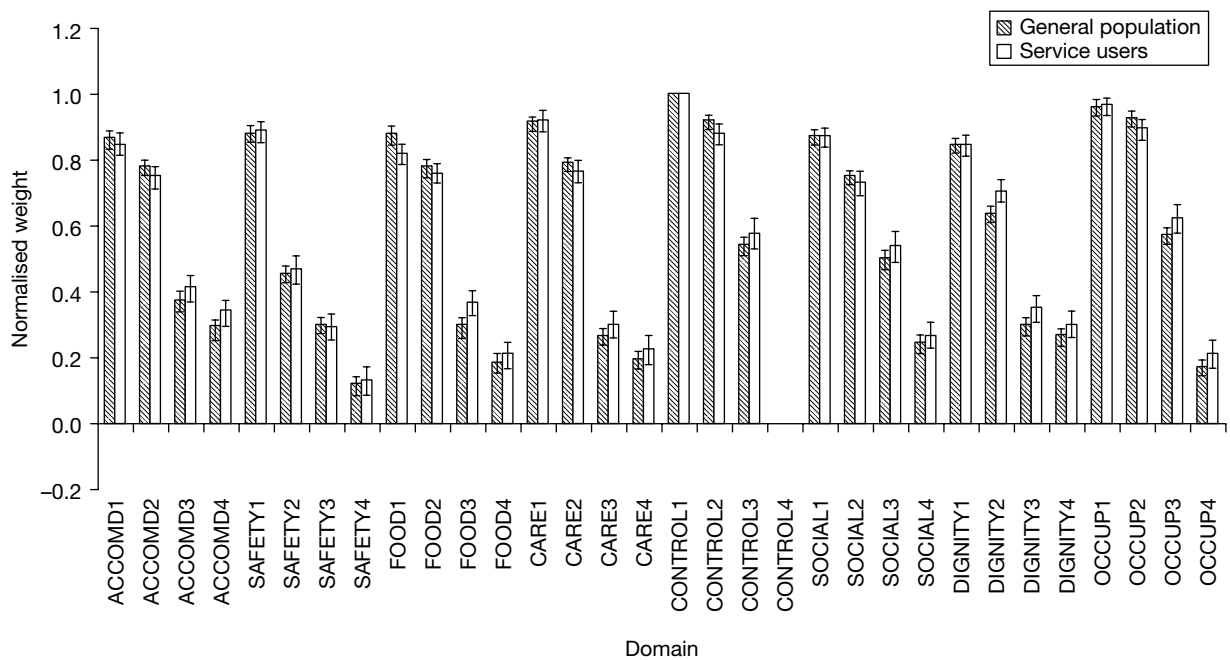
### Comparison of domain weights between the general population and service users

An important question to address was whether service users' preferences differed from those of the general population and, if so, whether or not these differences suggested alternative utility weights. We compared the results of the scale effects models, reported in *Tables 20* and *21*. These models had different scales, so the coefficients were not directly comparable.<sup>105</sup> However, we could look to the relative size of the differences using one of the domain levels as a common denominator and normalising all others relative to that. We chose the top level of the domain *control*, which was strongly estimated in both the general population and the service user models.

*Figure 21* shows the relative values and the 95% CIs (see Hess and Daly<sup>106</sup> for computation of SEs).



**FIGURE 20** Estimated weights and CIs in the general population limited scale effects model.



**FIGURE 21** Rescaled domain-level weights between the general population and service users.

We used *t*-tests to examine whether or not the normalised preference weights for each domain level differed significantly between the general population and service users. *Table 22* shows that preference weights were very similar among the general population and service users and in most cases were not statistically different. Only in 6 out of 30 domain-level weights were the valuations between the general population and service users significantly different at the 95% level.

Respondents in the general population sample placed higher value on attaining the top levels of *food and drink* ('I get all the food and drink I like when I want') and *control* ('I have as much control over my daily life as I want') than service users. On the other hand, service users placed higher value on the top level of *dignity* ('The way I'm helped and treated makes me think and feel better about myself'), compared with the general population. At the lower levels, it appeared that service users were slightly less concerned about being in a situation where:

- 'I don't always get adequate or timely food and drink' (*food and drink* at level 3),
- 'The way I'm helped and treated sometimes undermines the way I think and feel about myself' (*dignity* at level 3), or
- 'I do some of the things I value or enjoy with my time, but not enough' (*occupation and employment* at level 3).

However, none of the observed differences in value was large in absolute terms and in testing so many differences we would expect some differences to appear statistically significant by chance. Our conclusion was that there was insufficient evidence to warrant any separate estimates of utility weights to reflect the preferences of service users.

### **Domain weights: accounting for scale effects, position information and observable heterogeneity**

We also wanted to investigate what was associated with variation in preferences (taste heterogeneity). We examined the impact of:

- gender
- age
- level of education
- marital status
- number of children
- employment status
- household income
- type of area (urban, rural)
- place of residence
- ethnicity
- personal experience of social care (self)
- personal experience of social care (relatives).

A series of tests was run to compare the predicted probabilities of choosing each domain level against the observed frequencies of those being chosen within the data across each of these different respondent characteristics. Where these tests identified that the model led to a significant discrepancy between the observed and predicted choices, the specification of the utility functions was developed to take explicit account of the apparent differences in preferences. Taste heterogeneity was introduced into the model in two ways:

- additional dummy-coded variables on the utility of a given domain, representing a systematic difference in how identified groups valued the given domain relative to other domains (i.e. an effect that applies across all levels of that domain)

**TABLE 22** Comparison of estimated parameters between OSCA general population and service users

Domain level	General population		Service user		Significant at 95% CI
	Coefficient	SE	Coefficient	SE	
<b><i>Accommodation cleanliness and comfort</i></b>					
1. My home is as clean and comfortable as I want	0.86	0.014	0.85	0.017	No
2. My home is adequately clean and comfortable	0.78	0.013	0.75	0.018	No
3. My home is not quite clean or comfortable enough	0.37	0.016	0.41	0.021	No
4. My home is not at all clean or comfortable	0.29	0.016	0.34	0.021	No
<b><i>Safety</i></b>					
1. I feel as safe as I want	0.88	0.012	0.89	0.016	No
2. Generally I feel adequately safe, but not as safe as I would like	0.45	0.012	0.47	0.022	No
3. I feel less than adequately safe	0.30	0.012	0.29	0.022	No
4. I don't feel at all safe	0.11	0.013	0.13	0.022	No
<b><i>Food and drink</i></b>					
1. I get all the food and drink I like when I want	0.88	0.014	0.82	0.015	Yes
2. I get adequate food and drink at OK times	0.77	0.014	0.76	0.015	No
3. I don't always get adequate or timely food and drink	0.29	0.015	0.36	0.019	Yes
4. I don't always get adequate or timely food and drink, and I think there is a risk to my health	0.18	0.017	0.21	0.021	No
<b><i>Personal care</i></b>					
1. I feel clean and am able to present myself the way I like	0.91	0.011	0.92	0.016	No
2. I feel adequately clean and presentable	0.79	0.011	0.77	0.017	No
3. I feel less than adequately clean or presentable	0.27	0.013	0.30	0.021	No
4. I don't feel at all clean or presentable	0.19	0.013	0.23	0.022	No
<b><i>Control over daily life</i></b>					
1. I have as much control over my daily life as I want	1.00 (normalised level)				
2. I have adequate control over my daily life	0.92	0.010	0.88	0.016	Yes
3. I have some control over my daily life, but not enough	0.54	0.014	0.58	0.025	No
4. I have no control over my daily life	0.00 (base level)				
<b><i>Social participation and involvement</i></b>					
1. I have as much social contact as I want with people I like	0.87	0.010	0.87	0.014	No
2. I have adequate social contact with people	0.75	0.011	0.73	0.018	No
3. I have some social contact with people, but not enough	0.50	0.013	0.54	0.024	No
4. I have little social contact with people and feel socially isolated	0.24	0.012	0.27	0.022	No
<b><i>Dignity</i></b>					
1. The way I'm helped and treated makes me think and feel better about myself	0.85	0.011	0.85	0.016	No
2. The way I'm helped and treated does not affect the way I think or feel about myself	0.64	0.012	0.71	0.016	Yes
3. The way I'm helped and treated sometimes undermines the way I think and feel about myself	0.30	0.013	0.35	0.020	Yes
4. The way I'm helped and treated completely undermines the way I think and feel about myself	0.26	0.013	0.30	0.021	No
<b><i>Occupation and employment</i></b>					
1. I'm able to spend my time as I want, doing things I value or enjoy	0.96	0.011	0.97	0.014	No
2. I'm able do enough of the things I value or enjoy with my time	0.93	0.011	0.90	0.014	No
3. I do some of the things I value or enjoy with my time, but not enough	0.57	0.012	0.62	0.022	Yes
4. I don't do anything I value or enjoy with my time	0.17	0.012	0.21	0.022	No



- additional covariates on an individual domain level, capturing situations where identified groups placed a significantly higher (or lower) value on one particular domain level.

### General population

Table 23 shows the final model that captured both observable heterogeneity and scale effects for the general population. The goodness of fit of the model significantly improved, compared with the scale effects model ( $\chi^2_{(23df)} = 874 > \text{the critical value of } 49.7 \text{ at } 99.99\% \text{ confidence level}$ ). [Log-likelihood ratio test:  $\chi^2(\text{df observed heterogeneity model} - \text{df scale effects model}) = -2 \times (\text{LL}_{\text{obs. heterogeneity model}} - \text{LL}_{\text{scale effects model}})$ .]

The analysis revealed a number of associations with demographic and other characteristics likely to be related to people's expectations and experiences. The value placed on the top level of *food and drink* domain 'I get all the food and drink I like when I want' was lower for those with experience of knowing someone close who was unable to care for themselves. Respondents who self-rated themselves as having 'good' or 'alright' QoL placed a higher value on the top level of *personal care* than respondents who self-rated themselves as having 'bad', 'very bad' or 'so bad, it could not be worse' QoL. Also, those who lived in London, another city or a large town were less concerned about being in a situation where 'they would feel less than adequately clean or presentable' than respondents who lived in the suburbs of a city/large town, small town, a rural area or village.

London residents were also less concerned if they were in a situation in which they would 'have little social contact and feel socially isolated' than residents of another city/large town, suburbs of a city, a small town, or a rural area or village. Additional differences were observed in the valuation of the top and lowest levels of *dignity*. In particular, findings suggested that white British with an annual household income (before tax) of  $\geq \text{£}6500$  per year placed a higher value in being in a situation in which they would state 'The way I'm helped and treated makes me think and feel better about myself' than white British with an annual household income of  $< \text{£}6500$  or non-white respondents. Finally, residents in a London borough were less concerned about being in a situation describing the lowest level of *dignity* ('The way I'm helped and treated completely undermines the way I think and feel about myself').

Respondents in the general population sample who either received income support or income-based jobseeker's allowance placed a higher value on *accommodation cleanliness and comfort*. On the other hand, those who did not receive any type of benefit placed a higher value on the domain *food and drink*. Females, those who owned a house/flat outright or were buying a house/flat with a mortgage, and residents in London or the south-east of England placed a higher value on *safety* than the rest of the respondents in the general population sample.

The *control over daily life* domain was valued higher by those people who: were married (or equivalent status); received disabled person's tax benefit; owned a house/flat outright; had annual household income up to  $\text{£}50,000$ ; received council tax benefit; and did not have any children. In addition, London and South East residents valued higher the domain *social participation and involvement* than remaining respondents in the sample.

Finally, males, people without children, those who lived in suburbs of a city/large town, a small town, a rural area or village, or owned a house/flat outright or were buying a house/flat with a mortgage valued the domain *occupation and employment* higher.

### Service users

Service users' preferences were more closely associated with their own SCRQoL and experience of service. Table 24 shows the model, which again had significantly better goodness of fit

**TABLE 23** Estimated parameters using general population OSCA and MOPSU data with scale effects

Domain level	Coefficient	t-ratio (robust)
<b>Accommodation cleanliness and comfort</b>		
1. My home is as clean and comfortable as I want	4.52	27.1
2. My home is adequately clean and comfortable	4.01	25.6
3. My home is not quite clean or comfortable enough	1.58	14.4
4. My home is not at all clean or comfortable	1.14	11.1
1. My home is as clean and comfortable as I want (MOPSU)	4.21	25.6
2. My home is adequately clean and comfortable (MOPSU)	3.91	25.1
3. My home is less than adequately clean or comfortable (MOPSU)	1.68	14.6
4. My home is not at all clean or comfortable (MOPSU)	1.18	11.8
All levels: respondent received income support	0.45	3.1
All levels: respondent received income-based jobseeker's allowance	0.41	2.2
<b>Safety</b>		
1. I feel as safe as I want	4.60	26.6
2. Generally I feel adequately safe, but not as safe as I would like	2.04	19.2
3. I feel less than adequately safe	1.16	14.6
4. I don't feel at all safe – owns house/flat outright	0.55	4.7
4. I don't feel at all safe – does not own house/flat outright	0.00	n/a
All levels: respondent was female	0.26	3.7
All levels: respondent owned a house/flat outright or was buying house/flat with a mortgage	-0.3	-4.2
All levels: respondent lived in London or South East	0.27	3.8
<b>Food and drink</b>		
1. I get all the food and drink I like when I want* (with experience of knowing someone close unable to care for him/herself)	4.51	25.4
1. I get all the food and drink I like when I want* (without experience of knowing someone close unable to care for him/herself or not applicable)	5.06	24.8
2. I get adequate food and drink at OK times	4.08	24.5
3. I don't always get adequate or timely food and drink	1.05	10.5
4. I don't always get adequate or timely food and drink, and I think there is a risk to my health	0.40	4.2
1. I get all the food and drink I like when I want (MOPSU)	4.33	24.6
2. I get food and drink adequate for my needs (MOPSU)	4.11	24.4
3. I don't get all the food and drink I need, but I don't think there is a risk to my health (MOPSU)	1.45	12.8
4. I don't get all the food and drink I need, and I think there is a risk to my health (MOPSU)	0.58	5.9
All levels: individual did not receive any type of benefit	-0.14	-2.4
<b>Personal care</b>		
1. I feel clean and am able to present myself the way I like* (respondent's self-reported QoL was good or alright)	4.85	27.1
1. I feel clean and am able to present myself the way I like* (respondent's self-reported QoL was bad, very bad or so bad, it could not be worse)	4.00	11.5
2. I feel adequately clean and presentable	4.11	25.9
3. I feel less than adequately clean or presentable* (respondent lived in a London borough, another city/large town)	1.21	10.9
3. I feel less than adequately clean or presentable* (respondent lived in suburbs of a city/large town, a small town or a rural area or village)	0.81	8.8
4. I don't feel at all clean or presentable	0.55	7.7
<b>Control over daily life</b>		
1. I have as much control over my daily life as I want	4.75	26.5
2. I have adequate control over my daily life	4.32	26.3

**TABLE 23** Estimated parameters using general population OSCA and MOPSU data with scale effects (*continued*)

Domain level	Coefficient	t-ratio (robust)
3. I have some control over my daily life, but not enough	2.55	22.8
4. I have no control over my daily life	0.00	Base level
All levels: respondent was married or equivalent	0.25	3.7
All levels: respondent received disabled person's tax benefit	0.53	2.0
All levels: respondent owned house/flat outright or was buying house/flat with a mortgage	0.19	2.5
All levels: respondent had annual household income up to £50,000	0.20	3.2
All levels: respondent received council tax benefit	0.21	2.0
All levels: respondent did not have any children	0.30	4.7
<b><i>Social participation and involvement</i></b>		
1. I have as much social contact as I want with people I like	4.53	27.0
2. I have adequate social contact with people	3.80	25.5
3. I have some social contact with people, but not enough	2.32	19.2
4. I have little social contact with people and feel socially isolated* (respondent lived in a London borough)	1.40	10.3
4. I have little social contact with people and feel socially isolated* (respondent lived in another city/large town, suburbs of a city/large town, a small town or a rural area or village)	0.76	9.2
All levels: respondents lived London or South East	0.23	3.8
<b><i>Dignity</i></b>		
1. The way I'm helped and treated makes me think and feel better about myself* (white British with annual household income up to £6500 or did not know/refused to answer)	4.60	25.6
1. The way I'm helped and treated makes me think and feel better about myself* (white British with annual household income up to £6499 or non-white)	3.87	20.0
2. The way I'm helped and treated does not affect the way I think or feel about myself	3.15	21.8
3. The way I'm helped and treated sometimes undermines the way I think and feel about myself	1.12	12.8
4. The way I'm helped and treated completely undermines the way I think and feel about myself* (respondent lived in a London borough)	1.31	8.5
4. The way I'm helped and treated completely undermines the way I think and feel about myself* (respondent lived in another city/large town, suburbs of a city/large town, a small town or a rural area or village)	0.85	9.7
<b><i>Occupation and employment</i></b>		
1. I'm able to spend my time as I want, doing things I value or enjoy	4.67	26.1
2. I'm able do enough of the things I value or enjoy with my time	4.47	25.9
3. I do some of the things I value or enjoy with my time, but not enough	2.68	23.3
4. I don't do anything I value or enjoy with my time	0.81	11.0
All levels: respondent was male	0.21	3.5
All levels: respondent had no children	0.19	3.0
All levels: respondent lived in suburbs of a city/large town, a small town or a rural area or village	0.23	4.0
All levels: respondent owned a house/flat outright or was buying house/flat with a mortgage	0.18	3.0
<b><i>Domain position in the BWS task</i></b>		
Top: respondent chose the domain that appeared first	0.18	4.8
Top two: respondent chose the domain that appeared second	0.08	2.5
Bottom: respondent chose the attribute that appeared last	0.01	0.2
Bottom two: respondent chose the attribute that appeared before the last	-0.05	-1.5

*continued*

**TABLE 23** Estimated parameters using general population OSCA and MOPSU data with scale effects (*continued*)

Domain level	Coefficient	t-ratio (robust)
<b>Scale parameters</b>		
$\mu_{\text{excessive}}$ : MOPSU data with 'excessive randomisation'	0.84	3.5
$\mu_{\text{worst}}$ : data relate to a 'worst' or 'second-worst' choice	0.89	5.2
$\mu_{\text{not-imagine}}$ : data relate to individuals who could not put themselves in the imaginary positions described in the choices	0.80	3.1
$\mu_{\text{London and South-East}}$ : data relate to individuals living in London and South East	0.85	4.4
$\mu_{\text{health}}$ : data relate to individuals with self-reported fair, bad or very bad health	0.88	3.0
$\mu_{\text{int. duration}<15\text{min}}$ : data relate to individuals' interview duration shorter than 15 minutes	0.88	2.5
<b>Model diagnostics</b>		
No. of observations	29,728	29,728
df	72	72
Final log-likelihood	-41,890.0	-41,890.0
Rho <sup>2</sup> (0)	0.241	0.241

**TABLE 24** Estimated parameters using service user data with scale effects

Domain level	Coefficient	t-ratio (robust)
<b>Accommodation cleanliness and comfort</b>		
1. My home is as clean and comfortable as I want	5.84	13.7
2. My home is adequately clean and comfortable	4.99	13.0
3. My home is not quite clean or comfortable enough	1.91	8.1
4. My home is not at all clean or comfortable	1.41	7.0
All levels: respondent lived in south-east of England	-0.34	-2.5
<b>Safety</b>		
1. I feel as safe as I want	5.64	13.5
2. Generally I feel adequately safe, but not as safe as I would like	2.60	10.5
3. I feel less than adequately safe	1.43	7.3
4. I don't feel at all safe	0.30	1.8
All levels: respondent had 'adequate', 'some, but not enough' or 'little social contact with people and felt socially isolated'	0.57	4.4
<b>Food and drink</b>		
1. I get all the food and drink I like when I want	5.70	13.7
2. I get adequate food and drink at OK times	5.20	13.6
3. I don't always get adequate or timely food and drink	1.51	7.1
4. I don't always get adequate or timely food and drink, and I think there is a risk to my health* (respondent stated 'I have as much control over my daily life as I want' or 'I have adequate control over my daily life')	-0.06	-0.3
4. I don't always get adequate or timely food and drink, and I think there is a risk to my health* (respondent stated 'I have some control over my daily life, but not enough or 'I have no control over my daily life')	0.66	2.5
All levels: respondent was white British	-0.38	-3.3
<b>Personal care</b>		
1. I feel clean and am able to present myself the way I like	6.21	14.0
2. I feel adequately clean and presentable	4.99	12.8
3. I feel less than adequately clean or presentable	1.23	6.3
4. I don't feel at all clean or presentable* (respondent lived in south-east of England)	0.12	0.6
4. I don't feel at all clean or presentable* (respondent lived outside the south-east of England)	1.00	4.3

**TABLE 24** Estimated parameters using service user data with scale effects (*continued*)

Domain level	Coefficient	t-ratio (robust)
<b>Control over daily life</b>		
1. I have as much control over my daily life as I want – divorced or widowed/surviving partner	6.27	13.9
1. I have as much control over my daily life as I want – married (equivalent) or never been married (or equivalent) or separated (from spouse or partner) or refused/don't know	5.31	12.3
2. I have adequate control over my daily life	4.95	12.9
3. I have some control over my daily life, but not enough	3.15	11.9
4. I have no control over my daily life	0.00	Base level
All levels: respondent had not been to the day centre last month	0.45	2.7
All levels: respondent stated 'I have as much control over my daily life as I want' or (control) 'I have adequate control over my daily life'	0.27	2.1
All levels: respondent was white British	0.54	2.8
<b>Social participation and involvement</b>		
1. I have as much social contact as I want with people I like	5.55	13.6
2. I have adequate social contact with people	4.48	12.4
3. I have some social contact with people, but not enough	3.07	10.5
4. I have little social contact with people and feel socially isolated	1.19	6.4
All levels: respondent could do shopping by themselves	0.25	2.2
All levels: respondent stated that they had 'adequate', 'some social contact with people, but not enough' or had 'little social contact with people and they felt socially isolated'	0.30	2.7
<b>Dignity</b>		
1. The way I'm helped and treated makes me think and feel better about myself	5.78	13.2
2. The way I'm helped and treated does not affect the way I think or feel about myself	4.61	11.8
3. The way I'm helped and treated sometimes undermines the way I think and feel about myself	1.49	7.3
4. The way I'm helped and treated completely undermines the way I think and feel about myself	1.11	5.6
All levels: respondents stated that 'having help makes me think and feel better about myself' or 'Having help does not affect the way I think or feel about myself'	-0.26	-2.4
<b>Occupation and employment</b>		
1. I'm able to spend my time as I want, doing things I value or enjoy	5.90	14.0
2. I'm able do enough of the things I value or enjoy with my time	5.37	13.4
3. I do some of the things I value or enjoy with my time, but not enough	3.49	11.7
4. I don't do anything I value or enjoy with my time	1.07	6.2
All levels: respondent was white British	0.52	4.1
All levels: respondent owned house/flat outright	0.37	3.1
<b>Domain position in the BWS task</b>		
Top: respondent chose the domain that appeared first	0.35	4.6
Top two: respondent chose the domain that appeared second	0.16	2.5
Bottom: respondent chose the attribute that appeared last	0.00	0.1
Bottom two: respondent chose the attribute that appeared before the last	-0.07	-1.0
<b>Scale parameters</b>		
$\mu_{\text{worst}}$ : data relate to a 'worst' or 'second-worst' choice	0.74	8.0
$\mu_{\text{notSE}}$ : data relate to residents outside the south-east of England	0.66	8.5
$\mu_{\text{O-level}}$ : data relate to individuals with no formal qualification or just GCSE/O-level	0.84	2.9
<b>Model diagnostics</b>		
No. of observations	11,744	
df	52	
Final log-likelihood	-16,674.2	
Rho <sup>2</sup> (0)	0.235	

compared with the scale effects model ( $\chi^2_{(23df)} = 481 > \text{the critical value of } 36.1 \text{ at } 99.99\% \text{ confidence level}$ ).

Individuals who self-rated themselves as having low levels of control were less concerned than others about being in a situation where they would not always get adequate or timely food and drink. The domain *control* was valued higher by service users who did not use day care services, and who self-rated themselves as having 'adequate' or 'as much control over their daily life as they wanted'. Those who stated they could do shopping by themselves and respondents who had 'adequate', 'some social contact with people, but not enough' or had 'little social contact with people and they felt socially isolated' valued the domain *social participation and involvement* higher than the rest of the sample. Respondents who self-rated themselves as having 'adequate', 'some – but not enough' or 'little social contact with people and felt socially isolated' valued the domain *safety* higher than the rest of the sample. *Dignity* was valued lower by respondents who were experiencing adequate or high levels of dignity themselves in the help they received.

In terms of other characteristics, divorced or widowed respondents placed a higher value on the top level of the domain *control* ('I have as much control over my daily life as I want') than married (or equivalent) respondents or those who had never been married. Residents outside the south-east of England were less concerned about being in a situation in which they would state: 'I don't feel at all clean or presentable', and placed a lower value on *accommodation cleanliness and comfort*. White British service users placed a lower value on *food and drink* than non-white respondents, and, along with those who owned a house or flat outright, placed a higher value on the domain *occupation and employment*.

### Using the domain weights in application

We now turn to identifying the model on which we should base the utility weights to take forward for use with our measure in the ASCOT and other applications. The first question, whether or not we need separate weights to reflect service user preferences, has been resolved, as there is insufficient difference between their views and those of the general population to justify these. The weights based on the general population are based on a larger sample, so are more robust. Moreover, using population weights also corresponds closest with the practice followed in health care.

The second question relates to which model should be utilised: the one with just preference weights that takes account of the scale heterogeneity within the sample, or the one that takes account of both scale and preference heterogeneity. We decided to use the model with the scale heterogeneity, but without the preference heterogeneity. This model avoids the biases in the coefficient estimates that can result from scale heterogeneity, where the data from those groups that found the task harder would otherwise systematically bias the preference weights downwards, while providing weights that reflect the average valuation across the sample (which has been selected to be representative of the general population). In theory, the model that also includes the preference heterogeneity provides a fuller understanding of the population preferences. But, given that we are seeking a measure that is equally applicable to all within the population, we would need to take steps to provide a weighted average for each domain-level preference weight across all of the population, and this requires information on the proportions within each of the relevant segments.<sup>103</sup> This information is in many cases not known for the population as a whole, so the best estimates of these proportions are the proportions in our sample (which we know is broadly representative of the population in most key dimensions). In this case, there is little to be gained from weighting back to the sample, and the mean preferences from the model estimated across all respondents (once accounting for the scale heterogeneity) provide appropriate values for taking forwards.

We therefore focus on the models reported in *Table 20*, using the OSCA terms where alternative wording for domains was used in the two samples. In *Appendix 7* we also show equivalent models that have collapsed the four levels to three in each domain, as ASCOT includes a version with just three levels for use in care homes (see [www.pssru.ac.uk/ASCOT](http://www.pssru.ac.uk/ASCOT)).

In order to simplify the weights for application, we have further normalised the BWS coefficients to provide weights in the range 0–1, where level 4 of *control over daily life* = ‘0’ (worst level of all domains) and level 1 of *control over daily life* = ‘1’ (best level of all domains). These are shown in *Table 25*.

## Time trade-off study

We identified in *Chapter 4* that the aim of the TTO exercise was to enable us to adjust the domain-level weights (from the BWS data) to obtain QALY-equivalent weights for social care on a scale where ‘0’ represents ‘being dead’ and ‘1’ represents the social care equivalent of ‘perfect health’.

### Design

The design was based on the approach described in *Chapter 4*. The TTO task was undertaken using the computer-based implementation developed and piloted in the feasibility stage. In total, 64 different SCRQoL states were chosen to provide sufficient points to estimate a mapping function between BWS and mean TTO values. The states were selected from the full factorial of all possible states given the number of domains and their levels. The procedure for selecting these states was, first, to rank all states according to their utility value, which was computed based on the preference weights obtained in the BWS general population model. The states were defined as the threshold points which divided the utility scale into 64 equal intervals.

These 64 states were then compared with the ‘best SCRQoL’ state. Each respondent was given a randomly chosen subset of eight of these, drawing one state from each of eight ‘bins’ defined across the distribution to avoid clustering in the same part of the utility range.

### Data

A subsample of 126 respondents to the main general population BWS survey, who were willing to be interviewed again, were invited to undertake the TTO exercise. The target sample was 100 respondents. However, initial exploratory analysis of the data showed that a number of interviews were undertaken in very short time periods, and the TTO scores in these interviews followed a uniform distribution. Further analysis identified that interviews conducted by certain fieldworkers had short interview times. It was therefore decided by the project team to collect data from an additional sample of 26 respondents and exclude interviews with an unfeasibly short TTO task duration.

The sample had been randomly selected from those willing to be interviewed again, and there was no evidence of any real difference between this sample and the main BWS sample (see *Appendix 6*). In this follow-up interview, they were first asked to re-rate their current SCRQoL status, both to collect data for comparison with the earlier survey, and to re-familiarise the respondent with the domains. Interviewees were then introduced to the concept of a TTO task and taken through an example, after which they were asked to complete eight TTos themselves.

A number of exclusion criteria were applied to clean the data set prior to analysis. These exclusions removed anyone who:

**TABLE 25** Estimated parameters using general population with scale effects, normalised to 0–1

Domain level	Coefficient	Weight
<b><i>Accommodation cleanliness and comfort</i></b>		
1. My home is as clean and comfortable as I want	5.06	0.863
2. My home is adequately clean and comfortable	4.57	0.780
3. My home is not quite clean or comfortable enough	2.19	0.374
4. My home is not at all clean or comfortable	1.68	0.288
<b><i>Safety</i></b>		
1. I feel as safe as I want	5.16	0.880
2. Generally I feel adequately safe, but not as safe as I would like	2.65	0.452
3. I feel less than adequately safe	1.75	0.298
4. I don't feel at all safe	0.67	0.114
<b><i>Food and drink</i></b>		
1. I get all the food and drink I like when I want	5.15	0.879
2. I get adequate food and drink at OK times	4.54	0.775
3. I don't always get adequate or timely food and drink	1.72	0.294
4. I don't always get adequate or timely food and drink, and I think there is a risk to my health	1.08	0.184
<b><i>Personal care</i></b>		
1. I feel clean and am able to present myself the way I like	5.34	0.911
2. I feel adequately clean and presentable	4.62	0.789
3. I feel less than adequately clean or presentable	1.55	0.265
4. I don't feel at all clean or presentable	1.14	0.195
<b><i>Control over daily life</i></b>		
1. I have as much control over my daily life as I want	5.86	1.000
2. I have adequate control over my daily life	5.39	0.919
3. I have some control over my daily life, but not enough	3.17	0.541
4. I have no control over my daily life	0	0
<b><i>Social participation and involvement</i></b>		
1. I have as much social contact as I want with people I like	5.12	0.873
2. I have adequate social contact with people	4.38	0.748
3. I have some social contact with people, but not enough	2.91	0.497
4. I have little social contact with people and feel socially isolated	1.41	0.241
<b><i>Dignity</i></b>		
1. The way I'm helped and treated makes me think and feel better about myself	4.96	0.847
2. The way I'm helped and treated does not affect the way I think or feel about myself	3.73	0.637
3. The way I'm helped and treated sometimes undermines the way I think and feel about myself	1.73	0.295
4. The way I'm helped and treated completely undermines the way I think and feel about myself	1.54	0.263
<b><i>Occupation and employment</i></b>		
1. I'm able to spend my time as I want, doing things I value or enjoy	5.64	0.962
2. I'm able to do enough of the things I value or enjoy with my time	5.43	0.927
3. I do some of the things I value or enjoy with my time, but not enough	3.32	0.567
4. I don't do anything I value or enjoy with my time	1.00	0.170



- completed the TTO task in < 5 minutes
- scored each state equally at < 1.0 (we retained those who consistently scored all states as 1.0 as this is an acceptable response if the respondent is unwilling to distinguish between states and for ethical or religious reasons would not make a trade-off)
- consistently scored all states as worse than being dead
- scored the lowest state as having the highest score.

In total this led to the rejection of nine individuals, leaving 117 respondents for the TTO analysis.

## Analysis

The TTO score for each individual for each state was calculated using a scale of '0' ('being dead') to '1' ('all needs met'). For the cases where respondents valued states between these points, the usual linear TTO scoring scale was used to assign the score, i.e. the formula  $x/10$  where  $x$  is the number of years spent with 'all needs met'. For states that were rated as worse than being dead, the convention developed by Dolan<sup>112</sup> was adopted, where the score is given by the formula  $(x/10) - 1$ , bounding the values to  $-1$ .

From the TTO data, the mean TTO score for each of the 64 states was calculated to provide the average value assigned to each state. These average scores were used on the basis that we were seeking to understand the relationship between the TTO score and the corresponding BWS weight, with the latter calculated from models providing the average weights utility estimated across the sample. The average TTO score for a given state was then plotted against the corresponding weight for that state as implied by the BWS modelling, by summing across the domain levels that described the state in question (using BWS domain levels weights on a scale of 0–1 as reported in *Table 25*). This plot showed that the relationship between the mean TTO score and the BWS weight was linear, and some of the SCRQoL states were rated as worse than being dead (TTO score < 0).

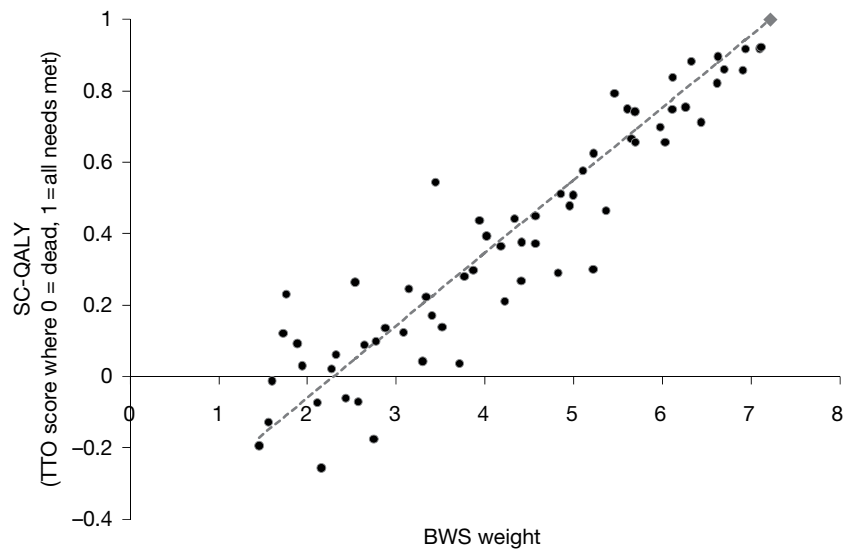
A simple mapping function form was estimated using

$$\text{OLS: } TTO_{ij} = f(DCE) + \varepsilon_{ij} \quad [\text{Equation 1}]$$

The first specification assumed a linear relationship with an intercept, and then squared and cubic terms were added to see whether or not performance is improved.

The regression was constrained to ensure that the 'all needs met' state had a value of 1.0 on the TTO scale, which is implied by how the TTO exercise is scored. The equation from this linear regression provides the transformation required to anchor the BWS weight to being dead, and provides a score that could be used to estimate a 'SC-QALY', which we have termed a SC-QALY index. The relationship between the two scoring schemes is shown in *Figure 22*. The results suggested that a simple linear relationship was sufficient, and indeed adding squared and cubic terms did not improve the relationship. There was a good overall fit of the model with an adjusted  $R^2$  of 0.859, a root mean squared error of 0.124 and mean absolute error across SCRQoL states of 0.091. It is worthy of note that this is the first occasion that we are aware of a TTO exercise being successfully applied to social care, and the extent of the linearity in the sample level preferences obtained from the TTO and BWS tasks suggests that the TTO may be usefully exploited to further understand the temporal dimension of the population preferences.

The regression between the TTO and BWS scores implies that the pit state, where all domains are at the high needs level, is valued at  $-0.171$  on the TTO scale. To give some context to this, the value that Dolan<sup>112</sup> estimated for the health EQ-5D pit state (using a comparable scoring scheme for states worse than being dead) was  $-0.543$ .



**FIGURE 22** Regression of mean TTO score against BWS weight for 64 SCRQoL states.  $SC-QALY = (0.203 \times BWS\_weight) - 0.466$ .

## Conclusion

The preference study work involved a number of different data collections from both the general population and service users. The results have shown a consistent picture in terms of both what we would hope to observe (higher values being put on higher SCRQoL states according to our definitions) and patterns of actual preferences. This has given us confidence in generating utility weights based on these analyses for our measure. In the next chapter, we turn to the final measure itself.

## Chapter 6

# The measure

### Introduction

The aim of the study was to develop a preference-weighted measure of social care outcomes that could be used for a wide variety of purposes. We built on previous work and took opportunities beyond this project to ensure that the measure was as well founded as possible. The previous chapters have described the methodological challenges and how these were met, both in developing the instrument and in identifying an anchored set of preference weights, and in the process investigating factors associated with these. In this final chapter, we provide an overview of the final measure, describing the distributions and demonstrating the application of the utility weights. We also call attention to some limitations of the study, discuss the potential uses and value of the measure, and identify future work needed.

### Population and service user item distributions

As part of the interviews undertaken to establish preferences of the general population and service users, respondents were asked about their current SCRQoL state in order both to familiarise them with the domains and levels, and to inform the analysis. This also provides us with useful comparative data on levels of SCRQoL in these two groups.

*Table 26* shows the responses to the individual items. In making comparisons, it should be borne in mind that service users were all people who had received equipment services, so their needs would predominantly be related to physical impairments rather than mental health problems. Moreover, they all were sufficiently cognitively able to undertake the BWS interview. As a result, their reported SCRQoL states are not necessarily representative of service users as a whole. The *dignity* question was asked of members of the general population only if they were currently receiving any help, which could be in the form of services or informal care. It was interesting that this applied to nearly one-quarter (24%) of the general population sample.

With the exception of *dignity*, members of the general population were significantly more likely to report the 'ideal' state and, with the exception of *occupation*, service users were significantly more likely to report the 'no needs' (i.e. where needs are met but not to the desired level), as well as 'needs' states, in each domain. This is to be expected if services are just meeting people's needs rather than further promoting capability, and suggests that the measure is sensitive to the distinction we were aiming to achieve by the additional level (see *Chapter 2*). The distribution of the *occupation* domain may reflect the nature of the general population sample which, although generally reasonably representative of the national population, had higher proportions of people who were retired (see *Appendix 6*). Unlike the other domains, the distribution of responses to *dignity* from members of the population who were receiving help was very similar to that of service users, although they were more likely to report that the way the help was provided made them feel better about themselves.

**TABLE 26** Distribution of responses to SCRQoL items in a general population and service user sample

Domain	General population (n=500)		Service user population (n=458)	
	Frequency	%	Frequency	%
<b><i>Control over daily life</i></b>				
I have as much control over my daily life as I want	309	61.8	128	28.0
I have adequate control over my daily life	139	27.8	156	34.1
I have some control over my daily life	45	9.0	145	31.7
I have no control over my daily life	7	1.4	29	6.3
<b><i>Personal cleanliness and comfort</i></b>				
I feel clean and am able to present myself the way I like	438	87.6	302	65.9
I feel adequately clean and presentable	54	10.8	136	29.7
I feel less than adequately clean or presentable	7	1.4	15	3.3
I don't feel at all clean or presentable	1	0.2	5	1.1
<b><i>Food and drink</i></b>				
I get all the food and drink I like when I want	449	89.8	334	72.9
I get adequate food and drink at OK times	37	7.4	107	23.4
I don't always get adequate or timely food and drink	12	2.4	16	3.5
I don't always get adequate or timely food and drink, and I think there is a risk to my health	2	0.4	1	0.2
<b><i>Accommodation cleanliness and comfort</i></b>				
My home is as clean and comfortable as I want	412	82.4	316	69.0
My home is adequately clean and comfortable	72	14.4	114	24.9
My home is not quite clean or comfortable enough	15	3.0	24	5.2
My home is not at all clean and comfortable	1	0.2	4	0.9
<b><i>Safety</i></b>				
I feel as safe as I want	336	67.2	183	40.0
I feel adequately safe, but not as safe as I'd like	149	29.8	167	36.5
I feel less than adequately safe	12	2.4	84	18.3
I don't feel at all safe	3	0.6	24	5.2
<b><i>Social participation</i></b>				
I have as much social contact as I want with people I like	328	65.6	168	36.7
I have adequate social contact	118	23.6	141	30.8
I have some social contact with people, but not enough	43	8.6	115	25.1
I have little social contact with people and feel socially isolated	11	2.2	34	7.4
<b><i>Occupation</i></b>				
I'm able to spend my time as I want, doing things I value or enjoy	246	49.2	103	22.5
I'm able to do enough of the things I value or enjoy with my time	147	29.4	124	27.1
I do some of the things I value/enjoy with my time, but not enough	100	20.0	192	41.9
I don't do anything I value or enjoy with my time	7	1.4	39	8.5
<b><i>Dignity<sup>a</sup></i></b>				
The way I'm helped and treated makes me think and feel better about myself	54	45.0	186	40.6
The way I'm helped and treated does not affect the way I think/feel about myself	39	32.5	159	34.7
The way I'm helped and treated sometimes undermines the way I think/feel about myself	23	19.2	91	19.9
The way I'm helped and treated completely undermines the way I think/feel about myself	4	3.3	22	4.8

a This question was applicable to (and responded to by) only 120 people in the general population sample.

## Utility weighting

Chapter 5 described the preference study analyses, providing a range of insights into the preferences of both the general population and service users with respect to SCRQoL. The principal objective, however, was to develop a utility weighting system that could be used for assessing different care interventions. It is useful, therefore, to review the steps required to apply the findings from the study to obtain the 'SC-QALY'.

The first step is to assess the SCRQoL state using the domain levels shown in Table 27.

For example, the individual may be assessed to have the following profile across the domains:

- *accommodation cleanliness and comfort* – level 2 'My home is adequately clean and comfortable'
- *safety* – level 3 'I feel less than adequately safe'
- *food and drink* – level 1 'I get all the food and drink I like when I want'
- *personal care* – level 2 'I feel adequately clean and presentable'
- *control over daily life* – level 3 'I have some control over my daily life, but not enough'
- *social participation and involvement* – level 4 'I have little social contact with people and feel socially isolated'
- *dignity* – level 3 'The way I'm helped and treated sometimes undermines the way I think and feel about myself'
- *employment and occupation* – level 2 'I'm able do enough of the things I value or enjoy with my time.'

This defines their SCRQoL state.

**TABLE 27** Domain level descriptions

Domain	Level 1	Level 2	Level 3	Level 4
Accommodation cleanliness and comfort	My home is as clean and comfortable as I want	My home is adequately clean and comfortable	My home is not quite clean or comfortable enough	My home is not at all clean or comfortable
Safety	I feel as safe as I want	Generally I feel adequately safe, but not as safe as I would like	I feel less than adequately safe	I don't feel at all safe
Food and drink	I get all the food and drink I like when I want	I get adequate food and drink at OK times	I don't always get adequate or timely food and drink	I don't always get adequate or timely food and drink, and I think there is a risk to my health
Personal care	I feel clean and am able to present myself the way I like	I feel adequately clean and presentable	I feel less than adequately clean or presentable	I don't feel at all clean or presentable
Control over daily life	I have as much control over my daily life as I want	I have adequate control over my daily life	I have some control over my daily life, but not enough	I have no control over my daily life
Social participation and involvement	I have as much social contact as I want with people I like	I have adequate social contact with people	I have some social contact with people, but not enough	I have little social contact with people and feel socially isolated
Dignity	The way I'm helped and treated makes me think and feel better about myself	The way I'm helped and treated does not affect the way I think or feel about myself	The way I'm helped and treated sometimes undermines the way I think and feel about myself	The way I'm helped and treated completely undermines the way I think and feel about myself
Employment and occupation	I'm able to spend my time as I want, doing things I value or enjoy	I'm able do enough of the things I value or enjoy with my time	I do some of the things I value or enjoy with my time, but not enough	I don't do anything I value or enjoy with my time

The state can then be scored by using the weights provided in *Table 28*, which, as *Chapter 5* showed, reflect the preferences of both the general population and service users.

In the case of this example, the domain level weights are:

- *accommodation cleanliness and comfort – level 2* domain level score = 0.780
- *safety – level 3* domain level score = 0.298
- *food and drink – level 1* domain level score = 0.879
- *personal care – level 2* domain level score = 0.789
- *control over daily life – level 3* domain level score = 0.541
- *social participation and involvement – level 4* domain level score = 0.241
- *dignity – level 3* domain level score = 0.295
- *employment and occupation – level 2* domain level score = 0.927.

This leads to the following calculation for the state score:

$$\text{State score} = 0.780 + 0.298 + 0.879 + 0.798 + 0.541 + 0.241 + 0.295 + 0.927 = 4.759 \quad [\text{Equation 2}]$$

This state score allows a comparison of the relative value placed on this state compared with other states, and for many applications will be sufficient. However, if desired, the state score can be translated into a score that could be used to estimate a 'SC-QALY' score, using the equation developed from the regression of TTO score against the BWS weight in *Chapter 5*:

$$\text{SC-QALY} = (0.203 \times \text{BWS\_weight}) - 0.466 \quad [\text{Equation 3}]$$

which in this case gives a state with the following SC-QALY score:

$$\text{SC-QALY} = (0.203 \times 4.759) - 0.466 = 0.500 \quad [\text{Equation 4}]$$

This provides an equivalent to the widely used measures of HRQoL, such as EQ-5D used as a basis for health QALYs. Although more work will be required to determine the extent to which this score can be used alongside or compared with the long-established QALY, the score does provide the analyst with the value associated with the given SCRQoL state on a scale where '0' = 'being dead' and '1' = 'best' or 'ideal' SCRQoL.

Where the interest is in establishing the 'expected' or 'gain' in SCRQoL, the same weights are applicable, as the questions used to establish 'expected' SCRQoL states use the same descriptors. Thus, in the example above, if an individual had high (level 4) needs for *personal cleanliness* and *food and drink* in the absence of services, his/her expected SCRQoL would be measured as:

$$\begin{aligned} \text{Expected state} &= 0.780 + 0.298 + 0.184 + 0.195 + 0.541 + 0.241 + 0.295 \\ &\quad + 0.927 = 3.461 \end{aligned} \quad [\text{Equation 5}]$$

$$\text{Expected SC-QALY} = (0.203 \times 3.461) - 0.466 = 0.237 \quad [\text{Equation 6}]$$

which can be deducted from the current score to provide a measure of SCRQoL gain:

$$\text{SC-QALY gain} = 0.500 - 0.237 = 0.263 \quad [\text{Equation 7}]$$

TABLE 28 Domain level weights

Domain level	Weight
<b><i>Accommodation cleanliness and comfort</i></b>	
1. My home is as clean and comfortable as I want	0.863
2. My home is adequately clean and comfortable	0.780
3. My home is not quite clean or comfortable enough	0.374
4. My home is not at all clean or comfortable	0.288
<b><i>Safety</i></b>	
1. I feel as safe as I want	0.880
2. Generally I feel adequately safe, but not as safe as I would like	0.452
3. I feel less than adequately safe	0.298
4. I don't feel at all safe	0.114
<b><i>Food and drink</i></b>	
1. I get all the food and drink I like when I want	0.879
2. I get adequate food and drink at OK times	0.775
3. I don't always get adequate or timely food and drink	0.294
4. I don't always get adequate or timely food and drink, and I think there is a risk to my health	0.184
<b><i>Personal care</i></b>	
1. I feel clean and am able to present myself the way I like	0.911
2. I feel adequately clean and presentable	0.789
3. I feel less than adequately clean or presentable	0.265
4. I don't feel at all clean or presentable	0.195
<b><i>Control over daily life</i></b>	
1. I have as much control over my daily life as I want	1.000
2. I have adequate control over my daily life	0.919
3. I have some control over my daily life, but not enough	0.541
4. I have no control over my daily life	0
<b><i>Social participation and involvement</i></b>	
1. I have as much social contact as I want with people I like	0.873
2. I have adequate social contact with people	0.748
3. I have some social contact with people, but not enough	0.497
4. I have little social contact with people and feel socially isolated	0.241
<b><i>Dignity</i></b>	
1. The way I'm helped and treated makes me think and feel better about myself	0.847
2. The way I'm helped and treated does not affect the way I think or feel about myself	0.637
3. The way I'm helped and treated sometimes undermines the way I think and feel about myself	0.295
4. The way I'm helped and treated completely undermines the way I think and feel about myself	0.263
<b><i>Employment and occupation</i></b>	
1. I'm able to spend my time as I want, doing things I value or enjoy	0.962
2. I'm able to do enough of the things I value or enjoy with my time	0.927
3. I do some of the things I value or enjoy with my time, but not enough	0.567
4. I don't do anything I value or enjoy with my time	0.170

## Population and service user distributions of the measure

For each SCRQoL state, three different summary scores can be produced:

- One score assumes equal weights for domains and response options and takes values from 0 to 24.
- Each level and domain can be weighted using the valuations from the BWS preference elicitation study to produce a current SCRQoL index, which can take values from 1.46 to 7.22.
- The SC-QALY index takes the BWS values and anchors these to the 'being dead' state at 0 as described above. This can take values from -0.17 to 1 (or 0.999 when rounded to three decimal places).

The distributions of these scores for both the equally weighted and preference-weighted current SCRQoL indices for the general population and service user population are shown in *Table 29*. *Figures 23* and *24* show the distributions of the SC-QALY index for these populations graphically. As might be hoped, no members of the population or service user group reported current SCRQoL states that would be rated as poorer than 'being dead'. Despite the compensatory activity of services, our measure suggests that SCRQoL is significantly better for the general population than for users of equipment services [using the SC-QALY index,  $F(1, 956) = 169.71, p < 0.001$ ].

## Validity and reliability

In *Chapter 3* we explored the construct validity of the descriptive system of the provisional measure with a sample of older home care service users. Although the preference study survey with users of equipment services was not designed specifically for exploring validity, the data collected did allow us to investigate many of the associations between sociodemographic, health and functional ability-related characteristics with a service user group that reflected a wider age range. Analysis of individual items that had been amended yielded a consistent pattern of results (see *Appendix 4*). A single item reporting QoL was significantly associated with the combined SCRQoL equally weighted and preference-weighted scores ( $R^2 = 0.57; p < 0.001$ ).

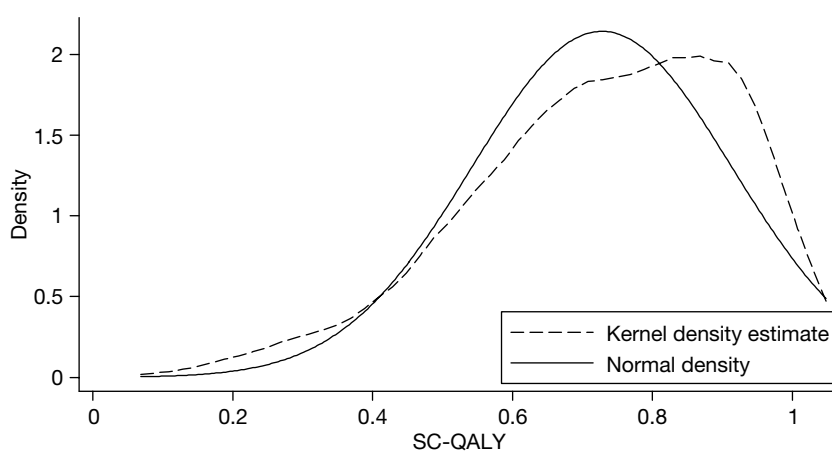
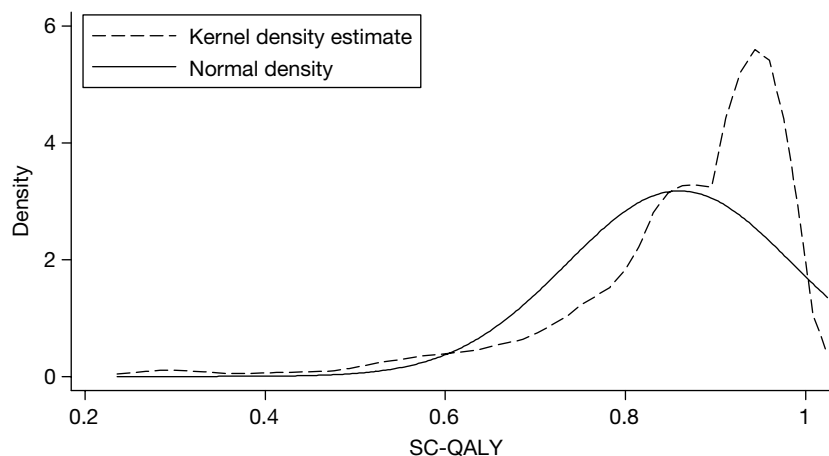
We investigated reliability through a follow-up survey of a subset of the equipment service users who participated in the BWS interview. Unfortunately, problems in conducting the fieldwork meant that the intended follow-up period of about 21 days was in practice over 74 days (SD = 18.6 days, max = 116 days, min = 25 days,  $n = 101$ ). We ran a variance components model to estimate the intraclass correlation coefficient (ICC) for the matched sample. The dependent variable was SCRQoL and the cluster variable the person. The reliability estimate was 0.68 when, ideally, we would want a reliability of  $> 0.8$ . The analysis was repeated splitting the sample between those who had the follow-up interview more or less than 60 days after the first interview. The ICC for the former was 0.65 and for the latter 0.75, implying that, had the guidance been followed and a shorter time period used, better test-retest reliability might have been observed. Under restricted maximum likelihood (REML) estimation for those interviewed within 60 days,  $n = 50$ , groups = 25; for those over 60 days,  $n = 150$ , groups = 75. The small number of groups for the shorter time period means that the estimates for the ICC may not be that accurate, but the results are indicative.

The indications were generally very positive with respect to the validity and reliability of the estimated preference weights. *Chapter 5* describes models that yielded consistent results, with higher utilities associated with higher levels of SCRQoL in each domain. Associations with



**TABLE 29** Distribution of equally weighted and preference-weighted versions of current SCRQoL index for general and service user populations

	Equally weighted		BWS preference weighted		SC-QALY index	
	General (n=500)	Service user (n=458)	General (n=500)	Service user (n=458)	General (n=500)	Service user (n=458)
Mean	20.48	17.58	6.53	5.88	0.86	0.73
SD	2.78	3.87	0.62	0.92	0.13	0.19
Max	24	24	7.22	7.22	1.00	1.00
Min	10	7	3.58	2.88	0.26	0.118
Skewness	-1.36	-0.31	-1.93	-0.64	-1.93	-0.64

**FIGURE 23** Distribution of the SC-QALY index for service user population sample (n=458). Kernel=Epanechnikov; bandwidth=0.0492.**FIGURE 24** Distribution of the SC-QALY index for general population sample (n=500). Kernel=Epanechnikov; bandwidth=0.0256.

preferences were logically consistent where we might have had prior expectations, and the results were remarkably stable. Very similar models were obtained from different general population samples taken a year apart and from service users in different geographical areas. The feasibility study demonstrated very similar results using DCE and BWS.<sup>105</sup> The preference weights on our final model using the TTO method were highly correlated with and demonstrated a strong

linear relationship with BWS estimates (see *Chapter 5*). Those factors found to be associated with variations in preferences were not ones (such as employment status) where our population sample differed from the general population. Thus, it is with some confidence that we can use the models to estimate our preference weights.

We have focused primarily on our measure of current SCRQoL state, but also noted that the compensatory nature of social care makes it particularly difficult to identify outcomes in practice. We have proposed a pragmatic approach that builds on this characteristic to allow individuals to identify the alternative, using the difference to generate a measure of SCRQoL 'gain'. The method is similar to the retrospective pre-test design proposed as a method for minimising the effects of response shift in other areas where such problems are encountered.<sup>96,113–116</sup> There is some evidence for the validity of the approach from this study (see *Chapter 3*), and with previous versions for day care<sup>7,9</sup> and care homes,<sup>8</sup> but more work is needed if it is to be widely applied.

## Limitations

Necessarily, there are a number of limitations to the study. In *Chapter 1* we described the nature of social care and identified the key role of the informal care network, with informal carers acting as both providers and beneficiaries of care. If we want to comprehensively reflect the impact of social care, we should be measuring outcomes for carers as well as for individuals with impairments. Our measure, although covering aspects of QoL that are relevant to everyone, reflects the objectives of services in relation to people with impairments.

As described in *Chapter 2*, our measure built on previous work on outcomes for older people and then developed to cover the range of domains relevant to younger age groups. Each domain covers a range of different aspects and it is feasible that there are other ways of describing the domains which might be more suited to specific groups. For example, a measure focused on younger adults with mental health problems might bring together more basic aspects of outcome, which are less frequently relevant to this group, and further divide domains relating to social participation and occupation.

However, the cognitive interviews did establish the applicability and relevance of all the domains to the full range of service user groups, confirmed by item response rates from different user groups,<sup>11</sup> and we were able to test the measure on users of equipment services, who include the full range of ages. Moreover, there were no indications that any domains were seen as irrelevant or any age-related variations in preferences. Although different SCRQoL measures might be better suited for certain groups, older people constitute the majority of social care service users, so it is appropriate that a generic measure ensures that the main areas of social care intervention relevant to this group are fully reflected.

When testing the measure, our service user samples were dependent on the UESs being conducted at the time of the study. This meant that much of the validity work was confined to older home care service users, and our service user survey samples were not ethnically diverse, although our population samples did have a good proportion of people from BME groups. Although we drew on data where they were available (e.g. from the IBSEN study<sup>49</sup>), for the most part it was not possible to involve people with cognitive impairments because of the cognitively demanding nature of much of the work.

We have investigated the descriptive validity of our items and the scales, and presented some evidence for the validity of the preference weights, but Brazier *et al.*<sup>70</sup> also identify empirical validity as a desirable characteristic of a preference-weighted measure. This has not been possible

within the scope of this study. Furthermore, although there is no evidence of lack of reliability, our attempts to provide robust evidence of reliability did not meet with success.

As argued in *Chapter 1*, we might expect adaptation to be a particularly relevant issue in social care. The way the SCRQoL measure is constructed would lead us to expect that it measures to some extent, and in so far as adaptation occurs, the adapted utility of service recipients. The effect of adaptation may be (partially) mitigated in two ways. First, in using general population preferences, the utility weights associated with different states will be much less affected by adaptation; general population respondents are considering hypothetical situations rather than actually experiencing limited capacity, and we expect their ability to predict their own adaptation to be limited. Second, the descriptive content of the instrument emphasises an aspiration outcome level rather than just 'usual' or normal levels of activity, and this may allow respondents to think beyond their own immediate situation or circumstances.

Nonetheless, we do anticipate that SCRQoL would be affected if people do adapt, and that effect is likely to lead to measuring a lower impact of services than if adaptation did not occur. It would be a useful development in the future to attempt to get some empirical estimate of the size of adaptation effects when using the instrument. While accepting that adaptation effects may occur, it is not at all clear whether, normatively, we should be measuring adapted utility or making allowances for adaptation where that is possible. The theoretical debate on this subject appears divided.<sup>44</sup> On the one hand, not adjusting for adaptation will lead to a lower priority being given to services and service recipients where adaptation occurs significantly compared with other service recipient groups when adaptation is absent or modest. On the other hand, making an adjustment implies that we give more weight to the utility of people who do adapt compared with the utility of those who do not. Also, how might this adjustment be made? It is not necessarily clear, therefore, that having an instrument that measures (potentially) adapted utility is a limitation. Nonetheless, it would clearly help in resolving this issue if we knew more about the scale and consequences of adaptation.

The preference weights estimated from the BWS data have utilised the S-MNL model, and this has been used to incorporate both sample-level variance heterogeneity and preference heterogeneity. The focus for this research has been on specifying models that take account of observable heterogeneity, but there are additional models now available that allow specifications that also incorporate unobserved heterogeneity, for example the generalised MNL (G-MNL) model and scale-adjusted latent class model. The estimation of these models was outside of the scope of this study as they were still the subject of academic investigation at the award of the research grant, but they have now started to be applied in practice and could be utilised in this context. The data collected here could be used to support the exploration of these alternative models, with a view to understanding whether or not the relative value placed on preference weights of different SCRQoL states are sensitive to the incorporation of unobserved heterogeneity in the model specification.

Arguably, all valuation studies should be stratifying factors associated with heterogeneity and variation in preferences, in order to ensure that the end weights accurately reflect the preferences of the relevant population. While we did not find systematic differences between our service user and general population sample preferences, service users' preferences were associated with their SCRQoL states. However, it is not feasible to establish the preferences of people in very poor SCRQoL states or people for whom the cause of impairment is associated with limited mental capacity, either in sufficient numbers or in some cases at all, for both practical and ethical reasons.

## Implications and potential future uses

As we describe in *Chapter 1*, the project was conducted alongside other related work, which has allowed the measure with provisional preference weights to be made available prior to the end of the project ([www.pssru.ac.uk/ASCOT](http://www.pssru.ac.uk/ASCOT)). The measure has attracted widespread national and international interest. In England, the interest is owing in part to its inclusion in the annual national ASCS. Both the overall measure and individual items form part of the Adult Social Care Outcomes framework.<sup>1</sup> This means that a valuable source of benchmark comparative data will be available from 2011. However, using the measure for this purpose suggests that further work on establishing a means of adjustment to reflect SCRQoL 'gain' would be of particular value in order that data on SCRQoL 'states' are not misinterpreted.

At present there are no QALY equivalent measures available in social care. This study has attempted to fill this gap. A SCRQoL-based SC-QALY would have many uses, paralleling the use of these indicators in health care. First, it is hoped that a SC-QALY will become a routine measure in the economic evaluation of social care interventions. Predecessors of the SCRQoL have already been used in Department of Health-funded evaluations of individual budgets in social care, reablement services, and in the evaluation of personal health budgets. The anchoring of the SCRQoL to create a SC-QALY will allow generic incremental cost-effectiveness comparisons to be made between social care interventions, and between social care and health-care interventions.

Second, a SCRQoL measure is useful information in social care practice. For example, it offers a means for regulators to ensure that providers are delivering minimum standards of SCRQoL to service recipients. Commissioners could use QoL information in informing commissioning choices and even reimbursement levels for providers. Currently, social care services are allocated on the basis of need (i.e. the severity of a person's condition and their non-service-related circumstances), but allocating resources on the basis of the potential marginal improvement in QoL offered by services is likely to lead to an overall improvement in productivity.

The study has identified that service user preferences are no different to those of the general population, but their experience of SCRQoL is. The study provides important insights into those preferences, supporting the policy emphasis on giving service users greater control. We have a basis on which to compare the SCRQoL of service users and the general population. Potentially, this provides us with a starting point for an 'outcomes' basis on which to debate and make difficult decisions about what we expect of social care services.

The anchoring exercise means that we now have estimated the necessary function for mapping BWS values onto TTO values. This means that, in theory, we can make direct comparison between ASCOT and EQ-5D values as they are on similar scales. Caution in interpretation will be needed as the upper anchor is different: EQ-5D is to 'full health', whereas ASCOT reflects the 'ideal' SCRQoL state. Nevertheless, ASCOT offers considerable potential to improve the sensitivity of economic evaluations to outcomes delivered, as health- and social care interventions are often closely linked, and to assist in making cost-utility comparisons across a wider range of intervention types than is currently possible.

## Further work

While the measure provides an important step forward, much remains to be done. The interdependent nature of well-being of carers and those they care for means that at present

the measure only partially reflects the impact of social care services. One preference-weighted measure has been developed,<sup>117-119</sup> but ideally we would want a measure that reflects more closely the SCRQoL measured by ASCOT. Some work has already been conducted exploring the applicability of the domains and adaptations needed to reflect the equivalent of SCRQoL for carers.<sup>120</sup> Although the results of this work has been promising,<sup>121</sup> there remain a number of challenges, not least establishing preference weights for this measure or a version of it, and identifying the best way to reflect trade-offs between the well-being of carers and cared-for individuals.

We have identified the limitations of applicability of the measure, as it stands, to those with cognitive impairment as the result of learning disabilities, dementia or other conditions. As part of the development work for the ASCS, a self-completion version for people with learning disabilities was developed, but it was acknowledged that this was not the ideal method for establishing SCRQoL for this group.<sup>8</sup> This was supported by a recent review of the ASCS.<sup>7</sup> A mixed-method approach for use in care homes has been developed.<sup>8</sup> However, this is resource intensive, and the high proportion of people with cognitive impairments of some sort who use adult social care services means further development work in this area would seem to be a priority.

While the findings are encouraging, evidence is needed about the reliability and sensitivity of the measure of current SCRQoL, facilitating correct interpretation of differences and change in people's experiences: when is a difference 'clinically significant'? Does context and type of impairment systematically affect responses?

As we note above, the use of ASCOT in the Adult Social Care Outcomes Framework (ASCOF) means that this will provide a valuable source of data, but if we are to reflect the 'value added' of services it is important that work is undertaken to aid interpretation. Further work building on the pragmatic approach reflected in the 'expected' and SCRQoL 'gain' measures could both feed into this and further test the validity and reliability of the approach.

Anchoring the measure to the 'being dead' state is only the first step if we are to be able to make best use of our SC-QALY alongside the long-established health care QALY. There are clearly relationships with EQ-5D, but many questions remain, not least of which is how our top anchor of 'ideal' SCRQoL state relates to 'full health'.

In light of these questions and considerations, in terms of future research we would recommend as priorities for future work:

- extending the approach to include informal carers
- exploring the potential to link between ASCOT and EQ-5D in estimating and comparing QALYs
- developing validated approaches to establishing ASCOT states for service users living in their own homes who have cognitive impairment and communication difficulties
- further work to test the validity, reliability and sensitivity of the current expected and gain measures in a variety of contexts.

These future developments will extend the range of outcomes that we are able to reflect, and the applicability and value of the ASCOT measure itself. Although this list can be seen as a rank order of priorities, they are largely independent of each other. Developments associated with the ASCS may provide opportunities to develop methods for establishing ASCOT states for a wider range of groups and further test reliability and validity. We put the highest emphasis on the carer measure as this reflects an additional aspect of outcome that ASCOT does not currently encompass.

## Conclusions

Resources are always limited, but this is even more the case in the current financial climate. This, together with the policy emphasis on outcomes, makes it all the more important that we have good measures to reflect the value of social care interventions for those who use them. In evaluating cost-effectiveness in health care, the QALY is widely accepted as a measure of outcome. Although related to health, social care has fundamentally different objectives, primarily compensating for rather than treating or mitigating impairment, meaning health QALYs will not accurately reflect the impact of social care services. The project was designed to establish the basis for an equivalent to the QALY for social care: a preference-weighted and ideally anchored generic measure of social care outcomes that could be used for a variety of purposes.

The first task was to clarify exactly what it is we are attempting to measure. Drawing on the SPOW and Sen's concepts of functioning and capabilities, we built on previous work to design a theoretically grounded measure of SCRQoL. In addition to a measure of current SCRQoL, which can be used to reflect differences over time and between different populations, we developed a pragmatic approach to establishing the impact of services, which draws on the ongoing nature of most social care, which means people themselves daily face the possibility that the help and support on which they rely may not be forthcoming. This means it is feasible to ask people about their needs in the absence of the help provided, allowing us to derive an 'expected' measure: the difference between people's current and expected SCRQoL reflecting the impact of services as SCRQoL 'gain'.

The items were cognitively tested, and the measure tested for validity and feasibility with social care service users. Response rates were good for all items, which implied that both the current and expected SCRQoL items are feasible with this population. Despite concerns about the distributions of two items, which led to further development and testing, there was good evidence for the validity of the descriptive system, each item capturing a different aspect of SCRQoL. There was also evidence for the validity of the SCRQoL current and gain scales, which appeared to operate as intended. The final version of the ASCOT instrument is shown in *Appendix 5*.

We explored different techniques in establishing preference weights and anchoring the scale, breaking new ground in a number of areas. We tested empirically the comparability of the profile-case BWS and DCE estimates with the general population. The similarity of the results and comparative advantage of BWS in terms of smaller cognitive burden, while viewing all the attributes and allowing the estimation of a common scale, led us to use this technique for the main preference elicitation task. We tested the feasibility of identifying user preferences and found that once the introduction and layout were clear, service users generally were able to undertake the tasks. We established an approach for identifying people for whom the task was too cognitively demanding, and ways of identifying and investigating the incidence and impact of strategies, such as taking into account only certain domains. In addition, we tested and then implemented the TTO technique, the first time this had been used with respect to social care, in order to enable us to anchor the scale to 'being dead', widening the ways in which the measure could be used and enhancing comparability with health QALYs.

The preference study work involved a number of different data collections and techniques. The results showed a consistent picture in terms of both what we would hope to observe (higher values being put on higher SCRQoL states according to our definitions) and patterns of actual preferences. An important question was whether or not service users' preferences differed from those of the general population. We found remarkably little difference. Indeed, the consistency in

the estimates generated across techniques, time, the general population and service users gave us considerable confidence in our final utility weights.

While there is clearly room for further work, the study has made important progress in this neglected area. The revised ASCOT measure developed through the study shows considerable potential, providing a first estimate of a social care equivalent to the QALY, and which can be used in a range of circumstances, including cost-effectiveness, cost-utility and policy evaluation, to reflect the impact and value of social care interventions. However, if we are to fully reflect the outcomes of social care, an equivalent measure is needed for carers. If we are to exploit the full potential of the measure, it will be important to establish how SC-QALYs relate to health QALYs and, although we have provided a starting point, there remains much to do in developing an approach that comprehensively reflects outcomes for all service users and exploring and interpreting differences between groups and circumstances. Our 'expected' and 'gain' measures build on the nature of social care and represent a radically new approach to reflecting outcome. While this would benefit from further work, we see this as a pragmatic low-cost approach that shows considerable promise.

There has been widespread interest since ASCOT was first made widely available in June 2010 through [www.pssru.ac.uk/ASCOT](http://www.pssru.ac.uk/ASCOT). The measure is already being incorporated in the annual national ASCS, and proposed as an important element of the Transparency in Social Outcomes framework, which should enhance its value by providing benchmark data about current SCRQoL states for a range of service user groups and contexts.





## Acknowledgements

We are very grateful to all of those who participated in the research: the LAs who took part in various stages and acted as 'critical friends'; Accent, who undertook the survey fieldwork; Diane Fox and James Caiels, who undertook cognitive interviews; Nick Smith, who provided data on the equipment survey; the Reference Group (Penny Banks, Bleddyn Davies, Caroline Glendinning, Martin Green, Mike Kelly, Mervyn Kohler, Jill Manthorpe, Kevin Marsh, Simon Medcalf, Lucianne Sawyer, Helen Tomkys) for their helpful input; and the Advisory Group (Kate Anderson, Mark Chandler, Mike Fisher, Tim Hind, Carol Jagger, Nicola Kay, Martin Knapp, Oliver Mills, Chris Rainey, Martin Stevens, Raphael Wittenberg) for their advice. Our thanks also go to Jane Dennett, Lesley Cox and Ann Richardson for their editorial and formatting input and skills, and to the anonymous referees for their helpful comments on the final report.

### Contributions of authors

Ann Netten (Director of PSSRU and Professor of Social Welfare) was principal investigator and project manager.

Peter Burge (Associate Director, Choice Modelling and Valuation) led the work on developing the preference weights.

Juliette Malley (Research Fellow) led the work on developing the items and testing validity, and conducted analysis associated with these activities.

Dimitris Potoglou (Analyst, Choice Modelling and Valuation) contributed to the design and analysis of the preference experiments.

Ann-Marie Towers (Research Officer) undertook validity testing analyses.

John Brazier (Professor of Health Economics) advised on the development of the descriptive system and the design and analysis of the TTO valuation.

Terry Flynn (Head of Social Policy & Economic Evaluation) developed the theoretical and analytical BWS models used to value preferences.

Julien Forder (Professor of the Economics of Social Policy) advised on the design of the outcome measure and its potential application in social care.

Beryl Wall (Associate Director, Market Research) managed the data collection activity and advised on the presentation of the TTO interview.

### Publication

1. Potoglou D, Burge P, Flynn T, Netten A, Malley J, Forder J, *et al.* Best-worst scaling vs. discrete choice experiments: an empirical comparison using social care data. *Soc Sci Med* 2011;**72**:1717–27.



## References

1. Department of Health (DoH). *Transparency in outcomes: a framework for adult social care*. London: DoH; 2010.
2. Department of Health (DoH). *Liberating the NHS: transparency in outcomes: a framework for the NHS*. London: DoH; 2010.
3. Atkinson A. *Atkinson review: final report. Measurement of government output and productivity for the National Accounts*. Basingstoke: Palgrave Macmillan; 2005.
4. Wanless D. *Securing good care for older people. Taking a long-term view*. London: King's Fund; 2006.
5. Netten A, Ryan M, Smith P, Skatun D, Healey A, Knapp M, *et al*. *The development of a measure of social care outcome for older people*. Discussion Paper 1690/2. Canterbury: PSSRU, University of Kent; 2002.
6. Netten A, Forder J, Shapiro J. *Measuring personal social services outputs for National Accounts: services for older people*. Discussion Paper 2267/3. Canterbury: PSSRU, University of Kent; 2006.
7. Office for National Statistics (ONS). *Measuring outcomes for public service users*. London: ONS; 2010.
8. Malley J, Caiels J, Fox D, McCarthy M, Smith N, Beadle-Brown J, *et al*. *A report on the development studies for the National Adult Social Care User Experience Survey*. Discussion Paper 2721. Canterbury: PSSRU, University of Kent; 2010.
9. Caiels J, Forder J, Malley J, Netten A, Windle K. *Measuring the outcome of low-level services: final report*. Discussion Paper 2699. Canterbury: PSSRU, University of Kent; 2010.
10. Burge P, Netten A, Gallo F. Estimating the value of social care. *J Health Econ* 2010;**29**:883–94.
11. Malley J, Netten A. *Putting people first: development of the putting people first user experience survey*. Discussion Paper 2637. Canterbury: PSSRU, University of Kent; 2009.
12. Netten A, Burge P, Malley J, Potoglou D, Brazier J, Flynn T, *et al*. *Outcomes of social care for adults (OSCA): interim findings*. Discussion Paper 2648/2. Canterbury: PSSRU, University of Kent; 2009.
13. NHS Information Centre for Health and Social Care. *Personal Social Services expenditure and unit costs England, 2008–09*. Leeds: NHS Information Centre for Health and Social Care; 2010.
14. Forder J, Towers A-M, Caiels J, Beadle-Brown J, Netten A. *Measuring outcomes in social care: second interim report*. Discussion Paper 2542. Canterbury: PSSRU, University of Kent; 2008.
15. Netten A. *Overview of outcome measurement in social care*. Canterbury: PSSRU, University of Kent; 2011.
16. Malley J, Towers A-M, Netten A. *Measuring social care related quality of life*. Discussion Paper 2751. Canterbury: PSSRU, University of Kent; 2011.
17. Knapp M. *The economics of social care*. Basingstoke: Macmillan; 1984.
18. Malley J, Fernández J-L. Measuring quality in social care services: theory and practice. *Ann Publ Cooper Econ* 2010;**81**:559–82.

19. Qureshi H, Nicholas E. A new conception of social care outcomes and its practical use in assessment with older people. *Res Pol Plan* 2001;**19**:11–26.
20. Department of Health (DoH). *Putting people first: a shared vision and commitment to the transformation of adult social care*. London: HM Government; 2009.
21. Windle K, Wagland R, Forder J, D'Amico F, Janssen D, Wistow G. *National evaluation of partnerships for older people projects: final report*. Discussion Paper 2700. Canterbury: PSSRU, University of Kent; 2009.
22. Birch S, Donaldson C. Valuing the benefits and costs of health care programmes: where's the 'extra' in extra-welfarism? *Soc Sci Med* 2003;**56**:1121–33.
23. Dolan P, Edlin R. Is it really possible to build a bridge between cost–benefit analysis and cost-effectiveness analysis? *J Health Econ* 2002;**21**:827–43.
24. Culyer A, editor. Rationing health care and the dilemmas it poses: some pale reflections on an agenda of Alan Williams. In *Be reasonable: following the William's Way*. York: Centre for Health Economics; 2006.
25. Hansen BO, Hougaard JL, Keiding H, Osterdal LP. On the possibility of a bridge between CBA and CEA: comments on a paper by Dolan and Edlin. *J Health Econ* 2004;**23**:887–98.
26. Department of Health (DoH). *Our health, our care, our say: a new direction for community services*, London: The Stationery Office; 2006.
27. Netten A, Davies B. The social production of welfare and consumption of social services. *J Publ Pol* 1990;**10**:331–47.
28. Netten A. The social production of welfare. In Knapp M, Fernández J-L, Challis D, Netten A, editors. *Towards equitable and efficient social care: festschrift for Bleddyn Davies*. Aldershot: Ashgate; 2004.
29. Becker G. *Human capital*. New York, NY: Columbia University Press; 1964.
30. Lancaster K. A new approach to consumer theory. *J Polit Econ* 1966;**74**:134–57.
31. Sen A. *Commodities and capabilities*. Amsterdam: North-Holland; 1985.
32. Sen A. The possibility of social choice. *Am Econ Rev* 1999;**89**:349–78.
33. Twigg J. Models of carers: how do social care agencies conceptualise their relationship with informal carers? *J Soc Pol* 1989;**18**:53–66.
34. van den Berg B, Bleichrodt H, Eeckhoudt L. The economic value of informal care: a study of informal caregivers' and patients' willingness to pay and willingness to accept for informal care. *Health Econ* 2005;**14**:363–76.
35. Department of Health (DoH). *Independence, well-being and choice: our vision for the future of social care for adults in England*. London: The Stationery Office; 2005.
36. Cabinet Office. *Improving the life chances of disabled people*. London: The Stationery Office; 2005.
37. Grewal I, Lewis J, Flynn T, Brown J, Bond J, Coast J. Developing attributes for a generic quality of life measure for older people: preferences or capabilities? *Soc Sci Med* 2006;**62**:1891.
38. Burchardt T. Capabilities and disability: the capabilities framework and the social model of disability. *Disabil Soc* 2004;**19**:735–52.
39. Oliver M. *Social work with disabled people*. Basingstoke: Macmillan; 1983.

40. Department of Health (DoH). *A vision for adult social care: capable communities and active citizens*. London: DoH; 2010.
41. Kahneman D, Diener E, Schwarz N. *Well-being: the foundations of hedonic psychology*. New York, NY: Russel Sage Foundation; 1999.
42. Kahneman D. Experienced Utility and objective happiness: a moment-based approach. In Kahneman D, Tversky A, editors. *Choices, values and frames*. New York, NY: Cambridge University Press; 2000.
43. Dolan P, Kahneman D. Interpretations of utility and their implications for the valuation of health. *Econ J* 2008;**118**:215–34.
44. Dolan P, Gudex C, Kind P, Williams A. The time trade-off method: results from a general population study. *Health Econ* 1996;**5**:141–54.
45. Williams A. *The measurement and valuation of health: a chronicle*. Discussion Paper 136. York: Centre for Health Economics, University of York; 1995.
46. Malley J, Sandhu S, Netten A. *Younger adults' understanding of questions for a service user experience survey: a report to the Information Centre for Health and Social Care*. Discussion Paper 2360. Canterbury: PSSRU, University of Kent; 2006.
47. Netten A, McDaid D, Fernández J-L, Forder J, Knapp M, Matosevic T, *et al*. *Measuring and understanding social services outputs*. Discussion Paper 2132/3. Canterbury: PSSRU, University of Kent; 2005.
48. Darton R, Forder J, Bebbington A, Netten A, Towers A-M, Williams J. *Analysis to support the development of FSS formulae for older people: final report*. Discussion Paper 2265/2. Canterbury: PSSRU, University of Kent; 2006.
49. Glendinning C, Challis D, Fernández J-L, Jacobs S, Jones K, Knapp M, *et al*. *Evaluation of the Individual Budgets Pilot Programme. Final report*. York: Social Policy Research Unit, University of York; 2008.
50. Bamford C, Bruce E. Defining the outcomes of community care: the perspectives of older people with dementia and their carers. *Ageing Soc* 2000;**20**:543–70.
51. Qureshi H, Patmore C, Nicholas E, Bamford C. *Outcomes of social care for older people and carers. Outcomes in Community Care Practice No. 5*. York: Social Policy Research Unit, University of York; 1998.
52. Qureshi H, Bamford C, Nicholas E, Patmore C, Harris J. Outcomes in social care practice: developing an outcome focus in care management and user surveys. Department of Health Report DH1738.6.00HQ. York: Social Policy Research Unit, University of York; 2000.
53. Shaping our Lives National User Network. *Shaping our lives: from outset to outcome. What people think of the social care services they use*. York: Joseph Rowntree Foundation; 2003.
54. Vernon A. *User-defined outcomes of community care for Asian disabled people*. Bristol: Policy Press; 2002.
55. Bamford C, Qureshi H, Nicholas E, Vernon A. *Outcomes of social care for disabled people and carers. Outcomes in Community Care Practice No. 6*. York: Social Policy Research Unit, University of York; 1999.
56. Dabelko-Schoeny H, King S. In their own words: participants' perceptions of the impact of adult day services. *J Gerontol Soc Work* 2010;**53**:176–92.
57. Raynes NV. Involving residents in quality specification. *Ageing Soc* 1998;**18**:65–78.

58. Train G, Nurock S, Kitchen G, Manela M, Livingston G. A qualitative study of the experiences of long-term care for residents with dementia, their relatives and staff. *Aging Ment Health* 2005;**9**:119–28.
59. Miller E, Cooper S-A, Cook A, Petch A. Outcomes important to people with intellectual disabilities. *J Pol Pract Intellect Disabil* 2008;**5**:150–8.
60. Henwood M, Lewis H, Waddington E. *Listening to users of domiciliary care services: developing and monitoring quality standards*. Leeds: Nuffield Institute for Health; 1998.
61. Raynes N, Temple B, Glenister C, Coulthard L. *Quality at home for older people; involving service users in defining home care specifications*. York: Joseph Rowntree Foundation; 2001.
62. Francis J, Netten A. Raising the quality of home care: a study of service users' views. *Soc Pol Admin* 2004;**38**:290–305.
63. Willis G, Lessler J. *Question appraisal system QAS-99*. Rockville, MD: Research Triangle Institute; 1999. URL: <http://appliedresearch.cancer.gov/areas/cognitive/guides.html> (accessed 15 May 2008).
64. Suchman L, Jordan B. interactional troubles in face-to-face survey interviews. *J Am Stat Assoc* 1990;**85**:232–41.
65. Willis G. *Cognitive interviewing: a tool for improving questionnaire design*. London: Sage; 2005.
66. Tourangeau R. Cognitive science and survey methods: a cognitive perspective. In Jabine TB, Straf ML, Tanur JM, Tourangeau R, editors. *Cognitive aspects of survey design: building a bridge between disciplines*. Washington, DC: National Academy Press; 1984. pp. 73–100.
67. Campanelli P, Martin E, Rothget J. The use of respondent and interviewer debriefing studies as a way to study the response error in survey data. *Statistician* 1991;**40**:253–64.
68. Brazier J, Ratcliffe J, Salomon J, Tsuchiya A. *Measuring and valuing health benefits for economic evaluation*. Oxford: Oxford University Press; 2007.
69. Messick S. *Validity of test interpretation and use*. Research Report No. 90–11. Princetown, NJ: Educational Testing Service; 1990.
70. Brazier J, Deverill M, Green C, Harper R, Booth A. A review of the use of health status measures in economic evaluation. *Health Technol Assess* 1999;**3**(9).
71. American Educational Research Association, American Psychological Association, National Council of Measurement in Education. *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association; 1999.
72. Lenert L, Kaplan RM. Validity and interpretation of preference-based measures of health-related quality of life. *Med Care* 2000;**38**:138–50.
73. Coast J, Peters TJ, Natarajan L, Sproston K, Flynn T. Valuing the ICECAP capability index for older people. *Soc Sci Med* 2008;**67**:874–82.
74. EuroQol group. EuroQol: a new facility for the measurement of health-related quality of life. *Health Pol* 1990;**16**:199–208.
75. Dolan P, Gudex C, Kind P, Williams A. Valuing health states: a comparison of methods. *J Health Econ* 1996;**15**:209–31.
76. Goldberg D, Hillier V. A scaled version of the General Health Questionnaire. *Psychol Med* 1979;**9**:139–45.

77. Goldberg D. Manual of the general health questionnaire. Windsor: National Foundation for Educational Research (NFER); 1978.
78. Goldberg D. *The detection of psychiatric illness by questionnaire*. London: Oxford University Press; 1972.
79. Hu Y, Stewart-Brown S, Twigg L, Weich S. Can the 12-item General Health Questionnaire be used to measure positive mental health? *Psychol Med* 2007;**37**:1005–13.
80. Hyde M, Wiggins RD, Higgs P, Blane DB. A measure of quality of life in early old age: the theory, development and properties of a needs satisfaction model (CASP-19). *Aging Ment Health* 2003;**7**:186–94.
81. Wiggins R, Netuveli G, Hyde M, Higgs P, Blane D. The evaluation of a self-enumerated scale of quality of life (CASP-19) in the context of research on ageing: a combination of exploratory and confirmatory approaches. *Soc Indic Res* 2008;**89**:61–77.
82. Wittenberg R, Comas-Herrera A, King D, Malley J, Pickard L, Darton R. *Future demand for long-term care, 2002 to 2041: projections of demand for long-term care for older people in England*. Discussion Paper 2330. London: London School of Economics; 2006.
83. Malley J, Towers A-M, Netten A, Brazier J, Forder J, Flynn T, *et al*. *An assessment of the construct validity of the ASCOT measure of social care-related quality of life*. Discussion Paper 2750. Canterbury: PSSRU, University of Kent; 2011.
84. Leyden KM. Social capital and the built environment: the importance of walkable neighbourhoods. *Am J Publ Health* 2003;**93**:1546–51.
85. Cohen J. *Statistical power analysis for the behavioral sciences*. 2nd edn. Hillsdale, NJ: Erlbaum; 1988.
86. Davies B, Fernández J-L, Nomer B. *Equity and efficiency policy in community care*. Aldershot: Ashgate; 2000.
87. Fernández J-L. *Utilisation and service productivities in community social care for older people: patterns and policy implications*. London: London School of Economics and Political Science; 2005.
88. Bleichrodt H. A new explanation for the difference between time trade-off utilities and standard gamble utilities. *Health Econ* 2002;**11**:447–56.
89. Louviere JJ, Street DJ, Burgess L, Wasi N, Islam T, Marley AAJ. Modelling the choices of single individuals by combining efficient choice experiment designs with extra preference information. *J Choice Modelling* 2008;**1**:128–63.
90. McFadden D. Conditional logit analysis of qualitative choice behavior. In Zarembka P, editor. *Frontiers in econometrics*. New York, NY: Academic Press; 1974. pp. 105–42.
91. Thurstone L. A law of comparative judgment. *Psychol Rev* 1927;**34**:273–86.
92. Louviere JJ, Hensher DA, Swait JD. *Stated choice methods: analysis and application*. Cambridge: Cambridge Press; 2000.
93. Fiebig DG, Keane M, Louviere JJ, Wasi N. The generalized multinomial logit model. *Marketing Sci* 2010;**29**:393–421.
94. Yatchew A, Griliches Z. Specification error in probit models. *Rev Econ Stat* 1985;**67**:134–9.
95. Herdman M, Badia X, Berra S. *EuroQol-5D: a simple alternative for measuring health-related quality of life in primary care*. Barcelona: Agencia d'Avaluacio de Tecnologia i Recerca Mediques; 2001.

96. Flynn TN, Louviere JJ, Peters TJ, Coast J. Using discrete choice experiments to understand preferences for quality of life. Variance scale heterogeneity matters. *Soc Sci Med*. 2010;**70**:1957–65.
97. Burge P, Gallo F, Netten A. *Valuing PSS outputs and quality changes*. Canterbury: PSSRU, University of Kent; 2006.
98. Ryan M, Netten A, Skatun D, Smith P. Using discrete choice experiments to estimate a preference-based measure of outcome: an application to social care for older people. *J Health Econ* 2006;**25**:927–44.
99. Louviere JJ, Woodworth GG. *Best-worst scaling: a model for largest difference judgments. Working Paper*. University of Alberta: Faculty of Business; 1990.
100. Finn A, Louviere JJ. Determining the appropriate response to evidence of public concern: the case of food safety. *J Publ Pol Marketing* 1992;**11**:12–25.
101. Marley A, Flynn TN, Louviere JJ. Probabilistic models of set-dependent and attribute-level best-worst choice. *J Math Psychol* 2008;**52**:281–96.
102. Flynn TN, Louviere JJ, Peters TJ, Coast J. Best-worst scaling: what it can do for health care research and how to do it. *J Health Econ* 2007;**26**:171–89.
103. Flynn TN. Using conjoint analysis and choice experiments to estimate quality adjusted life year values: issues to consider. *Pharmacoeconomics* 2010;**28**:711–22.
104. Potoglou D, Burge P, Flynn TN, Netten A, Malley J, Forder J, *et al*. Best-worst scaling vs. discrete choice experiments: an empirical comparison using social care data. *Soc Sci Med* 2011;**72**:1717–27.
105. Swait J, Louviere JJ. The role of the scale parameter in the estimation and comparison of multinomial logit models. *J Market Res* 1993;**30**:305–14.
106. Hess S, Daly A. *Calculating errors for measures derived from choice modelling estimates*. 88th Annual Meeting of the Transportation Research Board, Washington, DC, 2009.
107. Ratcliffe J, Brazier JE, Tsuchiya A, Symonds T, Brown M. Using DCE and ranking data to estimate cardinal values for health states for deriving a preference-based single index from the sexual quality of life questionnaire. *Health Econ* 2009;**18**:1261–76.
108. Islam T, Louviere JJ, Burke PF. Modelling the effects of including/excluding attributes in choice experiments on systematic and random components. *Int J Res Marketing* 2007;**24**:289–300.
109. Burge P, Potoglou D, Kim CW, Hess S. *How do the public value different outcomes of social care?* Cambridge: RAND Europe; 2010.
110. Flynn TN. Valuing citizen and patient preferences in health: recent developments in three types of best-worst scaling. *Expert Rev Pharmacoecon Outcomes Res* 2010;**10**:259–67.
111. Ben-Akiva M, Lerman SR. *Discrete choice analysis: theory and application to travel demand*. Cambridge: MIT Press; 1985.
112. Dolan P. Modelling valuations for EuroQol health states. *Med Care* 1997;**11**:1095–108.
113. Howard GS, Dailey PR. Response shift bias: a source of contamination of self-report measures. *J Appl Psychol* 1979;**64**:144–50.
114. Ring L, Hofer S, Heuston F, Harris D, O’Boyle C. Response shift masks the treatment impact on patient reported outcomes (PROs): the example of individual quality of life in edentulous patients. *Health Qual Life Outcomes* 2005;**3**:55.



115. Howard GS, Ralph KM, Gulanick NA, Maxwell SE, Nance SW, Gerber SK. Internal invalidity in pre-test-posttest self-report evaluations and a re-evaluation of retrospective pre-tests. *Appl Psychol Meas* 1979;**3**:1–23.
116. Sprangers M, Hoogstraten J. Pretesting effects in retrospective pretest-posttest designs. *J Appl Psychol* 1989;**74**:265–72.
117. Al-Janabi H, Coast J, Flynn TN. What do people value when they provide unpaid care for an older person? A meta-ethnography with interview follow up. *Soc Sci Med* 2008;**67**:111–121.
118. Al-Janabi H, Flynn TN, Coast J. Estimation of a preference-based carer experience scale. *Med Decis Making* 2011;**31**:458–68.
119. Al-Janabi H, Flynn TN, Coast J. QALYs and carers. *Pharmacoeconomics* 2011;**29**:1015–23.
120. Fox D, Holder J, Netten A. *Personal Social Services survey of adult carers in England: 2009–10: survey development project*. Discussion Paper 2643/2. Canterbury: PSSRU, University of Kent; 2010.
121. Malley J, Fox D, Netten A. *Developing a carers' experience performance indicator*. PSSRU Discussion Paper No. 2734. Canterbury: PSSRU, University of Kent; 2010.
122. Malley J, Sandhu B, Netten A. *Quality of social care for younger adults with physical and sensory impairments living in their own homes*. Discussion Paper No. 2770. Canterbury: PSSRU Research Unit, University of Kent; 2011.
123. McHorney CA, Cohen AS. Equating health status measures with item response theory. Illustrations with functional status items. *Med Care* 2000;**38**:II43–59.
124. De Vellis R. *Scale development: theory and applications*. 2nd edn. London: Sage; 2003.
125. Floyd FJ, Widaman KF. Factor analysis in the development and refinement of clinical assessment instruments. *Psychol Assess* 1995;**7**:286–99.
126. Olsson U. On the robustness of factor analysis against crude classification of the observations. *Multivariate Behav Res* 1979;**14**:485–500.
127. Holgado-Tello F, Chacón-Moscoso S, Barbero-García I, Vila-Abad E. Polychoric versus Pearson correlations in exploratory and confirmatory factor analysis of ordinal variables. *Qual Quant* 2010;**44**:153–66.
128. Olsson U. Maximum likelihood estimation of the polychoric correlation coefficient. *Psychometrika* 1979;**44**:443–60.
129. Fabrigar LR, Wegener DT, MacCallum RC, Strahan EJ. Evaluating the use of exploratory factor analysis in psychological research. *Psychol Meth* 1999;**4**:272–99.
130. Sijtsma K, Emons W, Bouwmeester S, Nyklíček I, Roorda L. Nonparametric IRT analysis of quality-of-life scales and its application to the World Health Organization Quality-of-Life Scale (WHOQOL-Bref). *Qual Life Res* 2008;**17**:275–90.
131. Moorer P, Suurmeijer TPBM, Foets M, Molenaar IW. Psychometric properties of the RAND-36 among three chronic disease (multiple sclerosis, rheumatic diseases and COPD) in the Netherlands. *Qual Life Res* 2001;**10**:637–45.
132. Sijtsma K, Molenaar IW, editors. *Introduction to nonparametric item response theory*. London: Sage; 2002.
133. Kaiser HF. An index of factorial simplicity. *Psychometrika* 1974;**39**:31–6.
134. Marley AAJ, Louviere JJ. Some probabilistic models of best, worst, and best-worst choices. *J Math Psychol* 2005;**49**:464–80.

135. Train K. *Discrete choice models with simulation*. Cambridge, UK: Cambridge University Press; 2003.
136. Ben-Akiva M, Morikawa T. Estimation of switching models from revealed preferences and stated intentions. *Transport Res* 1990;**24**:485–96.
137. ALOGIT. HCG Software, London, 2005. URL: <http://www.alogit.com>.
138. Bierlaire M, editor. BIOGEME: A free package for the estimation of discrete choice models. Proceedings of the 3rd Swiss Transportation Research Conference, Ascona, 2003.

## Appendix 1

### Development of domains of social care-related quality of life

OPUS (older people only)	OPUS (extension to other client groups)	ASCOT pre-OSCA	Final ASCOT
Food and nutrition	Meals and nutrition	Meals and nutrition	Food and drink
Personal care	Personal cleanliness and health	Personal hygiene	Personal cleanliness and comfort
Safety	Safety	Personal safety	Personal safety
Social participation and involvement	Social participation and involvement	Social participation and involvement	Social participation and involvement
Control over daily life	Control over daily life	Control over daily life	Control over daily life
	Accommodation cleanliness, order and accessibility	Accommodation cleanliness and comfort	Accommodation cleanliness and comfort
	Employment and occupation	Employment and occupation	Occupation <sup>a</sup>
	Having a caring role	Having a caring role	
	Living at home	Living at home	Dignity <sup>b</sup>
		Anxiety and worry <sup>c</sup>	
		Dignity <sup>a</sup>	

a Now encompassing caring role and employment and occupation.

b Redefined in relation to impact of care process on personal.

c Included only in a version developed for low-level services.<sup>9</sup>



## Appendix 2

# Analysis of previous versions of the Adult Social Care Outcome Toolkit

As we have described, there were a number of precursors to ASCOT, and versions of these instruments have been widely used in studies of social care interventions. Most recently, a version of the instrument was used in 2007 in a UES of adults aged 18–64 years with physical and sensory impairments, receiving help to enable them to live in their own homes, and in IBSEN. These two data sets were analysed to explore empirically, first, whether or not there are any problems with the items as currently phrased and, second, the choice of domains. We describe these two data sets before exploring both of these questions.

### Data sets

#### *The user experience survey of adults aged 18–64 years with physical and sensory impairments data set*

User experience surveys are postal self-completion surveys conducted annually by LAs across England. The LAs follow guidance on sampling, data collection and data management compiled by the Information Centre (IC) – the body charged with overseeing the survey and analysing the returned data – and send the data back to the IC as part of their annual data returns to central government. In 2007, the survey was for adults aged 18–64 years with physical and sensory impairments receiving help to enable them to live in their own homes.

For a number of years, PSSRU have conducted optional extension studies to the UES, whereby LAs can opt to use a longer version of the questionnaire. The longer questionnaire included questions on SCRQoL, as well as additional questions on quality of care and the characteristics of users and their services. Eleven councils across England agreed to take part in this optional PSSRU study. Further details about the sampling, data collection and management can be found in the report of this study.<sup>122</sup>

The users were on average about 50 years old ( $n = 2228$ , variance = 115.12), although the distribution was negatively skewed with median 54 years. Respondents were more likely to be female (59%) than male, the sample was predominantly white (90%), and a small but significant minority did not have English as a first language (6%).

Table 30 shows the disability and health status of the sample. As expected, service users rated their health as quite poor: compared with respondents to the 2006 Health Survey for England (HSE), the reported health of this sample was much worse. In the HSE, the majority of the general population rated their health as very good, and < 20% rated their health as fair or worse. The sample of service users was quite dependent on help to undertake activities of daily living, with the majority requiring help to dress and undress, transfer from bed or chair, and prepare a hot meal. Forty per cent of the sample also required help to wash their face and hands, a large proportion considering that needing help with this activity of daily living tends to indicate a high level of disability.<sup>123</sup>

**TABLE 30** Disability and health status of service users in the respondent sample ( $n = 2228$ )

Disability/health status	Frequency	%
<b>Functional limitations</b>		
Needs help to dress/undress ( $n = 1957$ )	1262	64.5
Needs help to transfer from bed or chair ( $n = 1838$ )	955	52.0
Needs help to wash face and hands ( $n = 1791$ )	704	39.3
Needs help to prepare hot meals ( $n = 1978$ )	1443	73.0
<b>Perceived health (<math>n = 2148</math>)</b>		
Very good	113	5.3
Good	443	20.6
Fair	935	43.5
Bad	459	21.4
Very bad	198	9.2

All members of the sample received a service from social services of at least one type (home care, direct payments, meals, day care or other services) and, on average, they had been receiving services for 4 years ( $n = 2200$ , variance = 22.08), although this distribution was positively skewed with a median of 2.4 years. For 23% of the sample, formal services were the only source of regular help. Of those who did receive regular practical help from friends, family or neighbours, around 45% received this help from someone living in their own household (including those receiving help from both outside and inside the home), leaving about 30% who received practical help on a regular basis from someone living outside their household.

### The individual budgets evaluation data set

The individual budgets evaluation (IBSEN) followed a randomised-control design and collected data from 959 individuals over 12 pilot sites in England by face-to-face interview. All the main service user groups were included in this study: older people (aged  $\geq 65$  years); younger PWPSI; younger PWLD; and younger PWMH. Further details about sampling, data collection and management can be found in the final evaluation report.<sup>49</sup> The pilots included people with a variety of impairments and difficulties, including PWLD, PWPSI, PWMH and older people with dementia and frailty. The interviews collected data on SCRQoL and a range of data about the needs characteristics of individuals, their health and disability, packages of care and QoL.

Key characteristics of the sample are summarised in *Table 31*. Analysis conducted as part of the IBSEN study suggested that the population of service users differed in some respects to the national service user population. The IBSEN sample contained more people receiving intensive services and more people from ethnic minorities than the general service user population.<sup>49</sup>

### Item analysis

Two sets of analysis were conducted:

- missingness analysis
- distribution of the items.

**TABLE 31** Average age, ethnicity and gender of the sample

Service user group	<i>n</i>	Mean age (years)	% female	% BME
Physical disability	327	54	63	10
Older people	263	81	66	5
Learning disability	235	34	42	11
Mental health	131	49	46	5
Total	956	57	56	8

Source: Evaluation of the Individual Budgets Pilot Programme Final Report.<sup>49</sup>

### **Analysis of the user experience survey for adults aged 18–64 years with physical and sensory impairments**

Across all of the SCRQoL items, at least one item was missing in 17% of cases. However, missingness varied significantly by the item, as is shown in *Table 32*. The *accommodation cleanliness and comfort* and *safety* domains had very high item response rates, whereas the *control over daily life* and *occupation* items had lower response rates, with 8% of cases having a missing response for the *occupation* item.

The distribution of responses to the SCRQoL items in the UES sample is shown in *Table 33*. For many of the domains (*control over daily life*, *personal cleanliness and comfort*, *food and nutrition* and *safety*), there are four response options where the first option is ‘needs met without help’, the second is ‘needs met with help’, the third is ‘low-level needs’ and the fourth is ‘high-level needs’. The ‘needs met with help’ level was introduced partly to meet the requests of LAs on the user survey group, who wanted to be able to identify where services were helping their clients. However, it was not possible to generate such a level for the *accommodation cleanliness and comfort*, *social participation* and *occupation* domains. In this analysis, we have combined the two ‘needs met’ categories for the relevant domains, as the existence of the expected SCRQoL questions preclude the need for the ‘needs met with help’ level. For most of the items, the majority of individuals responded to the top ‘needs met’ level. The distribution of the responses to the items is improved in the *accommodation cleanliness and comfort*, *social participation and enjoyment* and *occupation* domains where, in addition to the ‘needs met’ level, there is a response option that captures something slightly less than complete satisfaction.

### **Analysis of the individual budgets evaluation data set**

In the IBSEN data set, at least one item was missing in 12% of the cases. The pattern of missingness was similar across the items and client groups (*Table 34*). The items that contributed most towards the loss of data were the *social participation* and *occupation* domains, with rates of missingness at around 5–6%. Rates of missingness were higher for PWLD and PWMH across all of the items, and high for older people for the *social participation* and *occupation* domains. Rates were very low for PWPSI across all of the questions.

The distribution of responses to the SCRQoL items in the IBSEN sample is shown in *Table 35*. For most of the items, either the majority of individuals or the largest proportion responded to the top ‘needs met’ level, although for *control over daily life*, *safety* and *social participation* the respondents are more evenly distributed across the top two response options. For the two items with four levels (*accommodation cleanliness and comfort* and *occupation*), although in both cases the largest proportion responded to the top response option, the responses are relatively well distributed across all the response options. Very few people chose the lowest response option for the *personal cleanliness and comfort*, *food and nutrition* and *accommodation cleanliness*

**TABLE 32** Missingness statistics for responses for SCRQoL items ( $n = 2228$ )

Item	Item non-response	
	No.	%
Control over daily life	125	5.61
Personal cleanliness and comfort	101	4.54
Food and nutrition	108	4.85
Accommodation cleanliness and comfort	66	2.96
Safety	73	3.27
Social participation and enjoyment	97	4.36
Occupation	179	8.03

**TABLE 33** Distribution of responses to SCRQoL items in UES sample

Response option	Frequency	%
<b><i>Control over daily life</i></b>		
Feel in control/service help to feel in control	1531	72.8
Some control but not enough	453	21.5
No control	119	5.7
Total	2103	100.0
<b><i>Personal cleanliness and comfort</i></b>		
Feel clean and wear what want/with help	1895	89.1
Occasionally feel less clean	188	8.8
Feel much less clean than would like	44	2.1
Total	2127	100.0
<b><i>Food and nutrition</i></b>		
Eat meal like when want/with help	1768	83.4
Can't always eat meals like, but no health problems	264	12.5
Can't always eat and health risk	88	4.2
Total	2120	100.0
<b><i>Accommodation comfort and cleanliness</i></b>		
Home is clean and comfortable as like	993	45.9
Home is clean and comfortable as can be	766	35.4
Home could be more clean and comfortable	357	16.5
Home is not at all clean or comfortable	46	2.1
Total	2162	100.0
<b><i>Personal safety</i></b>		
No worries about safety/have support	1498	69.5
Some worries about safety	558	25.9
Extremely worried about safety	99	4.6
Total	2155	100.0



**TABLE 33** Distribution of responses to SCRQoL items in UES sample (*continued*)

Response option	Frequency	%
<b><i>Social participation</i></b>		
Good social life	261	12.2
Social life as good as can be	920	43.2
Have social life but sometimes feel lonely	504	23.7
Feel socially isolated and often lonely	446	20.9
Total	2131	100.0
<b><i>Occupation</i></b>		
Fully occupied in activities of my choice	987	48.2
Fully occupied not in activities of my choice	427	20.8
Don't have enough to do to keep me occupied	313	15.3
Nothing much to do and am usually bored	322	15.7
Total	2049	100.0

Totals vary for items because missing data is not reported.

**TABLE 34** Percentage of missing responses for SCRQoL items by client group (*n* = 1006)

Item	% cases missing item				
	Overall sample	Physical disability	Learning disability	Mental health	Older person
Control over daily life	2.88	0.59	5.39	4.38	2.93
Personal cleanliness	3.78	1.78	5.39	5.11	4.03
Food and nutrition	3.48	1.19	4.98	5.11	4.40
Accommodation	2.68	0.59	4.15	3.65	3.66
Safety	3.68	1.48	6.22	4.38	3.66
Social participation	5.86	2.97	7.88	5.11	8.42
Occupation	4.87	2.67	3.73	10.22	5.49
All items	12.13	6.23	14.94	16.06	15.02

*and comfort* domains. This is to be expected, as poor outcomes in these areas would present a serious risk to health, and we would expect social care services to be helping people to avoid such situations.

## Evidence for choice of domains

### Methods

To make choices between the attributes in valuation tasks, attributes must be orthogonal, or capture distinct concepts, a requirement referred to as 'preference independence' of the items.<sup>70</sup> Where measures are developed in a psychometric framework, multiple items are usually used to assess each attribute, and techniques – notably, exploratory factor analysis (EFA) – are used to assess the dimensionality of the items, the expectation being that the scale will be multidimensional, with as many dimensions as attributes. Where there is only one item per

**TABLE 35** Distribution of responses to SCRQoL items in IBSEN sample

Response option	Frequency	%
<b><i>Control over daily life</i></b>		
Feel in control	433	44.3
Some control but not enough	427	43.7
No control	117	12.0
Total	977	100.0
<b><i>Personal cleanliness and comfort</i></b>		
Feel clean and wear what want	778	80.4
Occasionally feel less clean	163	16.8
Feel much less clean than would like	27	2.8
Total	968	100.0
<b><i>Food and nutrition</i></b>		
Eat meal like when want	688	70.9
Can't always eat meals like, but no health problems	226	23.3
Can't always eat and health risk	57	5.9
Total	971	100.0
<b><i>Accommodation cleanliness and comfort</i></b>		
Home is clean and comfortable as like	595	60.8
Home is as clean and comfortable as it could be	237	24.2
Home could be more clean and comfortable	126	12.9
Home is not at all clean or comfortable	21	2.1
Total	979	100.0
<b><i>Safety</i></b>		
No worries about safety	446	46.0
Some worries about safety	415	42.8
Extremely worried about safety	108	11.1
Total	969	100.0
<b><i>Social participation</i></b>		
Good social life	345	36.4
Have social life but sometimes feel lonely	386	40.8
Feel socially isolated and often lonely	216	22.8
Total	947	100.0
<b><i>Occupation</i></b>		
Fully occupied in activities of my choice	433	45.2
I am occupied but not in activities of my choice	194	20.3
Don't have enough to do to keep me occupied	200	20.9
Nothing much to do and am usually bored	130	13.6
Total	957	100.0

Totals vary for items because missing data is not reported.

attribute, as is the case for ASCOT, the applicability of EFA is not obvious, as, if the attributes are orthogonal, there should not be any common variance among the items to explain. However, as the items measure the actual 'observed' SCRQoL state rather than the utility of that state, the attributes are actually likely to have common variance owing to the concerted action of various factors that influence observed SCRQoL, including the effect of underlying impairment and the action of social care.<sup>70,72</sup> It could be argued that if the concept of SCRQoL is valid and has meaning, we would expect the items to form a weak unidimensional scale, reflecting the common effect of social care on each attribute. Therefore, and counter to the psychometric position, the validity of the SCRQoL measure would be drawn into question only if the items have a very large degree of common variance and form a strong scale, as this would seem to indicate that items are substantively similar, rather than just affected by similar forces.

Factor models are commonly used to examine the structural relationship between questionnaire items, as the aim of the model is to explain the common variance of the items through a smaller number of latent variables, known as factors.<sup>124,125</sup> Prior to factor analysis, we examined the relationships between the items through examination of inter-item correlations. Polychoric correlations are used because of the ordinal nature of the measurement scale for each item.<sup>126</sup> [Polychoric correlations were calculated in Stata version 11 (StataCorp LP, College Station, TX, USA) using the user-written polychoric program.] The polychoric correlation matrix was used as the basis for exploratory factor analysis, as research has shown that Pearson correlations can lead to incorrect conclusions.<sup>127,128</sup> Exploratory analysis of both data sets revealed that the data were not multivariate normal. (*For the UES dataset:* Mardia's test for skewness = 21.496,  $\chi^2_{(84)} = 6677.313$ ,  $p < 0.001$ ; Mardia's test for kurtosis = 94.351,  $\chi^2_{(1)} = 3627.422$ ,  $p < 0.001$ ; Henze-Zirkler = 66.372,  $\chi^2_{(1)} = 47578.629$ ,  $p < 0.001$ ; Doornik-Hansen  $\chi^2_{(14)} = 13,230.371$ ,  $p < 0.001$ ; *for the IBSEN dataset:* Mardia's test for skewness = 9.198,  $\chi^2_{(84)} = 1360.885$ ,  $p < 0.001$ ; Mardia's test for kurtosis = 72.190,  $\chi^2_{(1)} = 148.066$ ,  $p < 0.001$ ; Henze-Zirkler = 13.800,  $\chi^2_{(1)} = 13,005.034$ ,  $p < 0.001$ ; Doornik-Hansen  $\chi^2_{(14)} = 2692.197$ ,  $p < 0.001$ .) We therefore used the principal axis factoring method for factor extraction, because this is most suitable when the assumption of multivariate normality does not hold.<sup>129</sup> Cases for which there were items with missing values were excluded from the analysis (17% of cases for the UES sample, and 12% for the IBSEN sample). To reflect the fact that the probability of endorsement of a response option varies across items in a graduated way, we also used a less familiar technique, known as Mokken scaling or non-parametric item response theory, to conduct the same analysis. We do not report this analysis here as the results are very similar to those from the more familiar factor analysis.

### Results from the adults aged 18–64 years with physical and sensory impairments survey data set

Table 36 shows the inter-item polychoric correlations for each of the SCRQoL items.<sup>130–132</sup> As in the distributional analysis, the 'needs met' and 'needs met with help' response options for the *control over daily life*, *personal cleanliness and comfort*, *food and nutrition* and *safety* domains have been combined. The correlations are all mostly moderate, being between 0.3 and 0.5.<sup>85</sup> However, some domain pairs have strong correlations, including *personal care* with the domains *control over daily life*, *food and nutrition* and *accommodation cleanliness and comfort*; *social participation* with the items *food and nutrition* and *occupation*; and *accommodation cleanliness and comfort* with the *safety* item. This could indicate that there is some conceptual overlap between these items, but it may also reflect a similarity in the concentration of service effort across these domains or users' preferences.

Tests for the suitability of the factor analysis procedure were all good. The Kaiser-Meyer-Olkin (KMO) test of sampling adequacy for this data set was 0.82, which is considered 'meritorious'.<sup>133</sup> Bartlett's test for sphericity rejected the null hypothesis that the variables are not intercorrelated [ $\chi^2_{(21)} = 2537.351$ ,  $p < 0.001$ ]. Therefore, we proceeded to factor analysis.

**TABLE 36** Matrix of polychoric correlations for SCRQoL items ( $n = 1860$ )

Item	Control	Personal cleanliness	Food	Accommodation	Safety	Social participation	Occupation
Control	1.000						
Personal	<b>0.541</b>	1.000					
Food	0.462	<b>0.592</b>	1.000				
Accommodation	0.382	<b>0.572</b>	0.462	1.000			
Safety	0.427	0.498	0.480	<b>0.503</b>	1.000		
Social	0.424	0.493	<b>0.503</b>	0.404	0.469	1.000	
Occupation	0.434	0.411	0.379	0.331	0.358	<b>0.628</b>	1.000

Correlations in bold text are  $>0.5$ .

Principal axis factoring was performed on the polychoric correlation matrix of the items ( $n = 1860$ ). A one-factor solution was obtained and is shown in *Table 37*. This factor has an eigenvalue of 3.24. The next factor had an eigenvalue of 0.2, which is well below the Joliffe (0.7) and Kaiser (1) criteria. All of the variables load on to the factor with relatively strong loadings. The unique variance is quite high for most of the variables, except *personal cleanliness and comfort*, and very high for the *occupation* item, which has a unique variance  $>0.6$ . This indicates that the factor does not explain this variable terribly well. Overall, the analysis implies that the items form a weak scale, as each item has a large unique variance.

### Results from the individual budgets evaluation data set

*Table 38* shows the inter-item polychoric correlations for each of the SCRQoL items. The correlations are all mostly moderate, being between 0.3 and 0.5, and some are weak ( $<0.3$ ).<sup>85</sup> However, some item pairs have strong correlations, including *personal care* with the items *food and nutrition* and *accommodation cleanliness and comfort*; and *social participation* with the *occupation* item. These pairs also had strong correlations in the UES sample.

Tests for the suitability of the factor analysis procedure were good. The KMO test of sampling adequacy for this data set was 0.79, which is considered 'middling'.<sup>133</sup> Bartlett's test for sphericity rejected the null hypothesis that the variables are not intercorrelated [ $\chi^2_{(21)} = 1030.194, p < 0.001$ ]. Therefore, we proceeded to factor analysis.

Principal axis factoring was performed on the polychoric correlation matrix of the items ( $n = 884$ ). A one-factor solution was obtained and is shown in *Table 39*. This factor has an eigenvalue of 2.71. The next factor had an eigenvalue of 0.32, which is well below the Joliffe (0.7) and Kaiser (1) criteria. All of the variables load onto the factor with good loadings above 0.4. However, the unique variance is quite high for most of the variables except *personal cleanliness and comfort* and very high for the *control over daily life, safety, accommodation cleanliness and comfort* and *food and nutrition* items, which all have unique variances of  $>0.6$ . This indicates that the factor does not explain these items variable terribly well. Overall, the analysis implies that the items form a weak scale, as many of the items have very high unique variances.

**TABLE 37** Pattern matrix for the one-factor solution ( $n = 1860$ )

Variable	Factor	Uniqueness
Control	0.642	0.588
Personal cleanliness	0.765	0.415
Food and nutrition	0.701	0.509
Accommodation	0.646	0.583
Safety	0.658	0.567
Social participation	0.718	0.484
Occupation	0.625	<b>0.609</b>

Items with uniqueness > 0.6 are shown in bold text.

**TABLE 38** Matrix of polychoric correlations for SCRQoL items ( $n = 884$ )

Item	Control	Personal cleanliness	Food	Accommodation	Safety	Social participation	Occupation
Control	1.000						
Personal	0.264	1.000					
Food	0.363	0.359	1.000				
Accommodation	0.207	0.295	<b>0.538</b>	1.000			
Safety	0.313	0.314	<b>0.563</b>	0.434	1.000		
Social	0.347	0.324	0.493	0.389	0.390	1.000	
Occupation	0.411	0.327	0.425	0.274	0.341	<b>0.617</b>	1.000

Correlations in bold text are > 0.5.

**TABLE 39** Pattern matrix for the one-factor solution ( $n = 884$ )

Variable	Factor	Uniqueness
Control	0.498	<b>0.752</b>
Safety	0.483	<b>0.767</b>
Personal cleanliness	0.747	0.443
Accommodation	0.581	<b>0.662</b>
Food and nutrition	0.632	0.600
Social participation	0.704	0.504
Occupation	0.658	0.567

Items with uniqueness > 0.6 shown in bold text.



## Appendix 3

### Meaning of response options by domain

#### Food and drink

##### Definition of domain

The service user feels that he/she has a nutritious, varied and culturally appropriate diet with enough food and drink that he/she enjoys at regular and timely intervals.

Level	Meaning
Preferred situation	Person receives a nutritious, varied and culturally appropriate diet, to the standards he/she wants
Needs met	Person receives a nutritious, varied and culturally appropriate diet (any deviations from this not being of concern to the individual), but he/she would prefer different food or at different times
Low-level need	The timing of meals is erratic, the person gets few (if any) hot meals, or the food is culturally inappropriate and causes the person some offence, although does not stop them eating it
High-level need	Person receives an inadequate diet that could potentially result in a health risk (would include when person does not eat the food because of cultural concerns)

#### Personal cleanliness and comfort

##### Definition of domain

The service user feels that he/she is personally clean and comfortable and looks presentable or, at best, is dressed and groomed in a way that reflects his/her personal preferences.

Level	Meaning
Preferred situation	Individual is personally clean, comfortable and feels he/she looks their best
Needs met	Individual is personally clean but feels he/she could look better
Low-level need	Person is not getting sufficient help with washing or dressing, and may be slightly embarrassed to see people because of this
High-level need	Person suffers from inadequate toileting or management of incontinence

#### Safety

##### Definition of domain

The service user feels safe and secure. This means being free from fear of abuse, falling or other physical harm, and free from fear of being attacked or robbed.

Level	Meaning
Preferred situation	Person feels safe and secure
Needs met	Person feels generally safe and secure and may have some worries, but these are not founded in any negative experiences
Low-level need	Person has worries that are caused by a bad experience, where the cause persists, for example a fall because of poor balance, crime in the neighbourhood and so on
High-level need	Person has worries that are caused by many negative experiences, such as continued abuse (physical or financial by staff) or physically inadequate buildings

## Social participation

### Definition of domain

The service user is content with his/her social situation, where social situation is taken to mean the sustenance of meaningful relationships with friends, family and feeling involved, or part of a community should this be important to the service user.

Level	Meaning
Preferred situation	Person is completely content with his/her relationships and involvement in the community
Needs met	Person is content with his/her relationships, but is not as involved and sociable as he/she would like to be
Low-level need	Person cannot participate in the community or keep up with friends and relatives as much as he/she would wish, and this has an impact on the way he/she feels about the support he/she receives from their relationships
High-level need	Person becomes socially isolated with little or no contact from others, and/or is distressed by his/her perceived level of isolation and lack of support

## Control over daily living

### Definition of domain

The service user can choose what to do and when to do it, having control over his/her daily life and activities.

Level	Meaning
Preferred situation	Person judges him/herself to have optimal control of their life
Needs met	Person judges him/herself to be in control of his/her life, but he/she is aware of limitations which are seen as inevitable (but might not be)
Low-level need	Person has some control and choice over clothing, bath time, bed times, etc., but these may be inconvenient and cannot be changed easily
High-level need	Person is in a situation where he/she has no choice over clothing, bath times, meals, meal times, going out, getting up and going to sleep, and so on

## Occupation

### Definition of domain

The service user is sufficiently occupied in a range of meaningful activities, whether formal employment, unpaid work, caring for others or leisure activities.

Level	Meaning
Preferred situation	Person is involved in many activities, and these are the types of things he/she enjoys doing
Needs met	Person is involved in many activities, which may not all be the activities he/she would choose to do or enjoys doing, but does feel sufficiently challenged and occupied
Low-level need	Individual is involved in some activities, but may feel quite bored a lot of the time and not mentally challenged
High-level need	Individual is not involved in any activities, or not involved in any activities he/she enjoys



## Accommodation cleanliness and comfort

### Definition of domain

The service user feels his/her home environment, including all the rooms, is clean and comfortable.

Level	Meaning
Preferred situation	The home is clean and the person feels completely comfortable
Needs met	Home is not kept to standards person would like, but he/she is not uncomfortable and the home is clean
Low-level need	Person does not feel comfortable in his/her home and is embarrassed when he/she invites people round. The home may be a bit dirty, particularly in rooms that he/she does not regularly/ever use
High-level need	Person feels he/she cannot have people coming round because all rooms are untidy and dirty, and the level of dirt is a potential risk to health

## Dignity

### Definition of domain

The negative and positive psychological impact of support and care on the service user's personal sense of significance.

Level	Meaning
Preferred situation	Services have a positive effect on self-esteem, person is unaware of being helped, and the helping process feels invisible
Needs met	Services do not have any effect on self-esteem, but person is aware of being helped and having to ask for assistance
Low-level need	Services have a slightly negative effect on person's self-esteem; he/she feels like a burden because of the actions of those who help him/her, for example rolling eyes
High-level need	Services have a serious negative effect on person's self-esteem; he/she feels invisible because of the way he/she is helped by others, for example talking over, being treated like an object not a person



## Appendix 4

# Validity of revised *accommodation* and *food and drink* items

The sample used for the following analysis was for 458 users of equipment services who participated in the preference study. The characteristics of the sample are shown in *Appendix 7*. *Table 40* shows the relationship between the measures of sociodemographic characteristics, general QoL and psychological well-being, health, control and autonomy, nature of the locality and environment, social contact and support, and participation and the two ASCOT attributes with new wording.

**TABLE 40** Test of associations between SCRQoL domains with the new wording (*food and drink*, and *accommodation cleanliness and comfort*) and other characteristics

Characteristics	p-value	
	Food and drink	Accommodation cleanliness and comfort
<b>Sociodemographic characteristics</b>		
Marital status ( <i>n</i> = 456)	0.264	<0.001***
Income ( <i>n</i> = 349) <sup>a</sup>	0.47	0.36
Tenure ( <i>n</i> = 458) <sup>a</sup>	0.70	0.57
Education ( <i>n</i> = 458)	0.31	0.09*
SEG ( <i>n</i> = 454)	0.39	0.10 <sup>a</sup>
<b>QoL</b>		
QoL ( <i>n</i> = 458) <sup>a,b</sup>	<0.001***	<0.001***
<b>Health</b>		
Self-perceived health ( <i>n</i> = 456) <sup>a,c</sup>	<0.001***	0.16
Outdoors and walk down road ( <i>n</i> = 456) <sup>a</sup>	0.10	0.011**
Get around indoors ( <i>n</i> = 456) <sup>a</sup>	<0.001***	0.001***
Get in/out bed ( <i>n</i> = 456) <sup>a</sup>	<0.001***	<0.001***
Bath, shower, wash all over ( <i>n</i> = 456) <sup>a</sup>	0.006**	<0.001***
Dressed/undressed ( <i>n</i> = 456) <sup>a</sup>	0.002***	<0.001***
Paperwork/finances ( <i>n</i> = 456) <sup>a</sup>	0.20	0.06**
Household shopping ( <i>n</i> = 456) <sup>a</sup>	0.005***	0.53
Prepare hot meals ( <i>n</i> = 449) <sup>a</sup>	0.13	0.40
<b>Nature of locality and environment</b>		
Living area ( <i>n</i> = 456) <sup>a</sup>	0.15	0.004***

SEG, socioeconomic group.

a Lowest two levels of the SCRQoL attribute are collapsed.

b Chi-squared used instead of Fisher's exact.

c Collapsed to only three levels (good, adequate and bad).

\*Significant at 10% level; \*\*significant at 5% level; \*\*\*significant at 1% level.

*n*= 458. Analysis used Fisher's exact test and across all response levels (unless otherwise indicated).

In the main, the relationships reported between the sociodemographic characteristics of this sample and the domains using the new wording were not significant. With the exception of marital status, the pattern of these associations is the same for equipment users as it is for older home care recipients.<sup>83</sup> Interestingly, marital status showed the opposite pattern of results to those found previously, being significantly, positively associated with *accommodation cleanliness and comfort* and not at all related to the *food and drink* domain. This may reflect differences between the two samples. This sample is younger and covers a more diverse range of service user groups. General QoL and all activities of daily living (ADL) except one (ability to go outdoors and walk down the road) showed strong and positive relationship with current SCRQoL in both domains. Compared with the wording used for the home care sample, the observed relationships with the ADL items were stronger. Generalising across the ADLs, the main effect tended to be that people who manage the ADLs on their own without difficulty are more likely to report being at the ideal state, and people managing on their own but with difficulty were more likely to be in the 'no needs' state. As we might expect with people receiving services, not being able to perform ADLs was not necessarily associated with needs (low/high).

In terms of instrumental activities of daily living (IADLs), undertaking household shopping but not the ability to prepare hot meals was significantly related to the *food and drink* domain. In the sample of older home care recipients, we found the opposite. In this sample, people who do their own shopping are more likely to be in the ideal state for *food and drink*, and people who rely on others because they cannot do it for themselves are more likely to be 'no needs' or 'needs' (high/low collapsed). People who do not currently shop for themselves because someone else does it for them (e.g. in terms of role distribution within the household) rather than because they cannot manage it themselves are distributed across the needs levels as expected, showing no particular relationship. This indicates that our measure is picking up exactly what we intended it to, capability ('I can do it or could do so if I wanted to') and functioning ('I can't do it and have to rely on others'). It seems that, in this sample, services were compensating better for needs in the *food and drink* domain because the significant relationship between ability to prepare hot meals and current SCRQoL in the home care sample disappeared ( $p=0.13$ ).

Finally, living area showed similar associations with our domains, as we had previously found (a significant relationship with *accommodation cleanliness and comfort* but not *food and drink*), but the pattern of results was slightly different. For older home care recipients, we found that people living in London and large towns were better off, being more likely to report the ideal state for *accommodation cleanliness and comfort*. For this sample, however, we found the opposite was true, with people living in rural areas being more likely to report the ideal state, people in small towns being more likely to report the 'no needs' state, and people in London and other large towns being slightly more likely to report needs. This might be a function of differences in the participating councils and the nature of the sample.

## **Appendix 5**

# **Final version of the Adult Social Care Outcome Toolkit instrument**

**Note: this document has not been re-formatted according to HTA conventions.**

To help us to measure the impact of services and support we are talking to the people who actually use them. We think people themselves are best able to judge what things would be like if services or support were not there. The next set of questions may seem strange to you but we are trying to get your views on the impact of services.

When we talk about services and support from Social Services in the next set of questions we mean for you to think about <<...>>

*Interviewer note: Insert an automated statement that comes from the set of questions about services.*

**1. Could you tell me which of the following statements best describes how much control you have over your daily life?**

*If needs a prompt then please say: By 'control over daily life' we mean having the choice to do things or have things done for you as you like and when you want*

I have as much control over my daily life as I want

I have adequate control over my daily life

I have some control over my daily life but not enough

I have no control over my daily life

**2. Do the support and services that you get from Social Services help you to maintain control over your daily life?**

*NOTE to interviewer: if the person needs prompting or reminding please say by support and services that you get from Social Services, we mean <<...>> (filled in as above)*

Yes

No

Don't know

If 2=yes, then

- 3. **Imagine that you didn't have the support and services from Social Services that you do now and no other help stepped in. In that situation, which of the following would best describe the amount of control you'd have over your daily life?**

NOTE to interviewer: if the person needs prompting or reminding please say by support and services that you get from Social Services, we mean <<....>> (filled in as above). It is important that people do not base their answers on the assumption that any help steps in, please emphasise this to interviewees.

(Reassure if necessary: please be assured that this is purely imaginary and does not affect the services you receive in any way.)

I would have as much control over my daily life as I want

I would have adequate control over my daily life

I would have some control over my daily life but not enough

I would have no control over my daily life

- 4. **Thinking about your personal care, by which we mean being clean and presentable in appearance, which of the following statements best describes your situation?**

I feel clean and am able to present myself the way I like

I feel adequately clean and presentable

I feel less than adequately clean or presentable

I don't feel at all clean or presentable

- 5. **Do the support and services that you get from Social Services help you to stay clean and presentable?**

NOTE to interviewer: if the person needs prompting or reminding please say by support and services that you get from Social Services, we mean <<....>> (filled in as above)

Yes

No

Don't know

If 5=yes, then

6. **Imagine that you didn't have the support and services from Social Services that you do now and no other help stepped in. Which of the following would then best describe your situation with regard to your personal care?**

*NOTE to interviewer: if the person needs prompting or reminding please say by support and services that you get from Social Services, we mean <<....>> (filled in as above)*

*It is important that people do not base their answers on the assumption that any help steps in, please emphasise this to interviewees.*

*(Reassure if necessary: please be assured that this is purely imaginary and does not affect the services you receive in any way.)*

I would feel clean and would be able to present myself the way I like

I would feel adequately clean and presentable

I would feel less than adequately clean or presentable

I wouldn't feel at all clean or presentable

7. **Thinking about the food and drink you have, which of the following statements best describes your situation?**

I get all the food and drink I like when I want

I get food and drink adequate for my needs

I don't get all the food and drink I need, but I don't think there is a risk to my health

I don't get all the food and drink I need, and I think there is a risk to my health

8. **Do the support and services that you get from Social Services help you to get the food and drink you want or need?**

*NOTE to interviewer: if the person needs prompting or reminding please say by support and services that you get from Social Services, we mean <<....>> (filled in as above)*

Yes

No

Don't know



If 8= yes then

9. **Imagine that you didn't have the support and services from Social Services that you do now and no other help stepped in. Which of the following would then best describe your situation with regard to food and drink?**

*NOTE to interviewer: if the person needs prompting or reminding please say by support and services that you get from Social Services, we mean <<....>> (filled in as above)*

*It is important that people do not base their answers on the assumption that any help steps in, please emphasise this to interviewees.*

*(Reassure if necessary: please be assured that this is purely imaginary and does not affect the services you receive in any way.)*

I would get all the food and drink I like when I want

I would get food and drink adequate for my needs

I wouldn't get all the food and drink I need, but I don't think there would be a risk to my health

I wouldn't get all the food and drink I need, and I think there would be a risk to my health

10. **Could you tell me which of the following statements best describes how clean and comfortable your home is?**

My home is as clean and comfortable as I want

My home is adequately clean and comfortable

My home is less than adequately clean or comfortable

My home is not at all clean or comfortable

11. **Do the support and services that you get from Social Services help you to keep your home clean and comfortable?**

*NOTE to interviewer: if the person needs prompting or reminding please say by support and services that you get from Social Services, we mean <<....>> (filled in as above)*

Yes

No

Don't know

If 11= yes then

12. **Imagine that you didn't have the support and services from Social Services that you do now and no other help stepped in. In that situation, which of the following would best describe how clean and comfortable your home is?**

*NOTE to interviewer: if the person needs prompting or reminding please say by support and services that you get from Social Services, we mean <<....>> (filled in as above)*

*It is important that people do not base their answers on the assumption that any help steps in, please emphasise this to interviewees.*

*(Reassure if necessary: please be assured that this is purely imaginary and does not affect the services you receive in any way.)*

- My home would be as clean and comfortable as I want
- My home would be adequately clean and comfortable
- My home would be less than adequately clean or comfortable
- My home would not be at all clean or comfortable

13. **Could you tell me which of the following statements best describes how safe you feel?**

*By feeling safe we mean feeling safe both inside and outside the home. This includes fear of abuse, falling or other physical harm and fear of being attacked or robbed*

- I feel as safe as I want
- Generally I feel adequately safe, but not as safe as I would like
- I feel less than adequately safe
- I don't feel at all safe

14. **Do the support and services that you get from Social Services help you to feel safe?**

*NOTE to interviewer: if the person needs prompting or reminding please say by support and services that you get from Social Services, we mean <<....>> (filled in as above)*

- Yes
- No
- Don't know

If 14 = yes then

15. **Imagine that you didn't have the support and services from Social Services that you do now and no other help stepped in. In that situation, which of the following would best describe how safe you feel?**

*NOTE to interviewer: if the person needs prompting or reminding please say by support and services that you get from Social Services, we mean <<....>> (filled in as above)*

*It is important that people do not base their answers on the assumption that any help steps in, please emphasise this to interviewees.*

*(Reassure if necessary: please be assured that this is purely imaginary and does not affect the services you receive in any way.)*

I would feel as safe as I want

Generally I would feel adequately safe, but not as safe as I would like

I would feel less than adequately safe

I wouldn't feel at all safe

16. **Thinking about how much contact you've had with people you like, which of the following statements best describes your social situation?**

I have as much social contact as I want with people I like

I have adequate social contact with people

I have some social contact with people, but not enough

I have little social contact with people and feel socially isolated

17. **Do the support and services that you get from Social Services help you to have contact with people you like?**

*NOTE to interviewer: if the person needs prompting or reminding please say by support and services that you get from Social Services, we mean <<....>> (filled in as above)*

Yes

No

Don't know

If 17 = yes then

18. **Imagine that you didn't have the support and services from Social Services that you do now and no other help stepped in. In that situation, which of the following would best describe how much contact you have with people you like?**

*NOTE to interviewer: if the person needs prompting or reminding please say by support and services that you get from Social Services, we mean <<....>> (filled in as above)  
It is important that people do not base their answers on the assumption that any help steps in, please emphasise this to interviewees.  
(Reassure if necessary: please be assured that this is purely imaginary and does not affect the services you receive in any way.)*

I would have as much social contact as I want with people I like

I would have adequate social contact with people

I would have some social contact with people, but not enough

I would have little social contact with people and would feel socially isolated

19. **Could you tell me which of the following statements best describes how you spend your time?**

*If respondent needs prompting please say: When you are thinking about how you spend your time, please include anything you value or enjoy including formal employment, voluntary or unpaid work, caring for others and leisure activities.*

I'm able to spend my time as I want, doing things I value or enjoy

I'm able do enough of the things I value or enjoy with my time

I do some of the things I value or enjoy with my time but not enough

I don't do anything I value or enjoy with my time

20. **Do the support and services that you get from Social Services help you to spend your time doing things you value and enjoy?**

*NOTE to interviewer: if the person needs prompting or reminding please say by support and services that you get from Social Services, we mean <<....>> (filled in as above)*

Yes

No

Don't know

If 20 = yes then

21. **Imagine that you didn't have the support and services from Social Services that you do now and no other help stepped in. In that situation, which of the following would best describe how you spend your time? Please assume that any other help you currently have would remain the same.**

*NOTE to interviewer: if the person needs prompting or reminding please say by support and services that you get from Social Services, we mean <<...>> (filled in as above)*

*It is important that people do not base their answers on the assumption that any help steps in, please emphasise this to interviewees.*

*(Reassure if necessary: please be assured that this is purely imaginary and does not affect the services you receive in any way.)*

I would be able to spend my time as I want, doing things I value or enjoy

I wouldn't be able to do enough of the things I value or enjoy with my time

I would do some of the things I value or enjoy with my time but not enough

I wouldn't do anything I value or enjoy with my time

22. **Which of these statements best describes how having help to do things makes you think about feel about yourself?**

Having help makes me think and feel better about myself

Having help does not affect the way I think or feel about myself

Having help sometimes undermines the way I think and feel about myself

Having help completely undermines the way I think and feel about myself

23. **Thinking about the way you are helped and treated, and how that makes you think and feel about yourself, which of these statements best describes your situation?**

The way I'm helped and treated makes me think and feel better about myself

The way I'm helped and treated does not affect the way I think or feel about myself

The way I'm helped and treated sometimes undermines the way I think and feel about myself

The way I'm helped and treated completely undermines the way I think and feel about myself



## Appendix 6

# Main best–worst scaling method preference study and follow-up samples

### General population main sample

The OSCA general population survey was undertaken with 500 respondents located in Birmingham, Newcastle, London and the south-east of England. Some of the key characteristics of respondents, compared with the MOPSU and general population as reported in the 2001 Census, are provided in *Tables 41–52*.

*Table 41* shows that the distribution of males and females was consistent across OSCA, MOPSU and the 2001 Census.

*Table 42* shows there was good distribution across age categories in our samples. It is notable that the survey samples slightly over-represent those aged 50–64 years and those aged 65–79 years, but under-represent those aged  $\geq 80$  years. The under-representation of this oldest age band is

**TABLE 41** Gender distribution

Gender	OSCA		MOPSU		Total		Population (%)
	Frequency	%	Frequency	%	Frequency	%	
Male	245	49	241	47.0	486	48	48.6
Female	255	51	272	53.0	527	52	51.3
<b>Total</b>	<b>500</b>	<b>100</b>	<b>513</b>	<b>100</b>	<b>1013</b>		<b>100</b>

[Population: Office for National Statistics mid-2008 estimate, table 4 (England)].

**TABLE 42** Age distribution

Age (years)	OSCA		MOPSU		Total		Population (%)
	Frequency	%	Frequency	%	Frequency	%	
18–24	57	11.4	57	11.5	114	11.3	12.1
25–30	54	10.8	61	11.9	115	11.4	10.1
31–39	89	17.8	70	13.6	159	15.7	15.8
40–49	78	15.6	98	19.1	176	17.4	18.8
50–64	121	24.2	132	25.7	253	25.0	22.8
65–69	37	7.4	84	16.4	165	16.3	14.7
70–79	44	8.8					
80–89	19	3.8	11	2.1	31	3.1	5.8
$\geq 90$	1	0.2					
<b>Total</b>	<b>500</b>	<b>100</b>	<b>513</b>	<b>100</b>	<b>1013</b>	<b>100</b>	<b>100</b>

[Population: Office for National Statistics mid-2008 estimate, table 4 (England)].

not entirely surprising as these are harder-to-reach individuals. The distribution of the pooled samples is broadly representative of the general population.

*Table 43* shows that, compared with the 2001 Census data, respondents were slightly more likely than the general population to have no formal education or to have post-school qualifications.

As shown in *Table 44*, only nine out of 1013 respondents belonged to the highest SEG. More than half of the respondents belonged to band C. Again, the distribution is broadly consistent across the two surveys. When compared with the general population, the OSCA sample includes a greater proportion of individuals in band C2 or higher and lower proportions of people in bands D and E.

*Table 45* shows that in both survey samples the proportion of respondents that reported being divorced are notably higher than would be expected from the Census data. However, the samples appear to be otherwise broadly representative.

*Table 46* shows that the samples under-represent both those working full time and those classifying themselves as full- or part-time students. The samples significantly over-represent those who are retired. This latter observation is interesting, given that the age distribution of

**TABLE 43** Level of education

Education	OSCA		MOPSU		Total		Population (%)
	Frequency	%	Frequency	%	Frequency	%	
No formal qualification	128	25.6	123	24.0	251	24.8	18.9
GCSE/O-level	148	29.6	152	29.6	300	29.6	34.4
A-levels or equivalent	75	15	52	10.1	127	12.5	12.4
Professional qualification below degree level	75	15	93	18.1	168	16.6	8.7
Degree level qualification or equivalent	52	10.4	65	12.7	117	11.6	7.8
Higher degree	12	2.4	16	3.1	28	2.8	3.5
Other	10	2	12	2.3	22	2.2	14.4
<b>Total</b>	<b>500</b>	<b>100</b>	<b>513</b>	<b>100</b>	<b>1013</b>	<b>100</b>	<b>100</b>

[Population: 2001 Census, National report for England and Wales, table S115 (England and Wales)].

**TABLE 44** Socioeconomic status

Band	OSCA		MOPSU		Total		Population (%)
	Frequency	%	Frequency	%	Frequency	%	
A	5	1	4	0.8	9	0.9	22.0
B	123	24.6	111	21.6	234	23.1	
C1	145	29	141	27.5	286	28.2	29.7
C2	96	19.2	126	24.6	222	22	15.1
D	59	11.8	68	13.3	127	12.5	17.2
E	72	14.4	61	11.9	133	13.1	16.1
Not stated			2	0.4	2	0.2	–
<b>Total</b>	<b>500</b>	<b>100</b>	<b>513</b>	<b>100</b>	<b>1013</b>	<b>100</b>	<b>100</b>

[Population: 2001 Census, National report for England and Wales, table S066 (England and Wales)].



TABLE 45 Marital status

Status	OSCA		MOPSU		Total		Population (%)	
	Frequency	%	Frequency	%	Frequency	%		
Married, or equivalent	285	57	Married	248	48.3	597	58.9	51.6
			Living together	64	12.5			9.9
Never been married, or equivalent	111	22.2	Single	101	19.7	212	20.9	23.1
Separated (from spouse or partner)	18	3.6	Separated	11	2.1	29	2.9	1.9
Divorced	44	8.8	Divorced	42	8.2	86	8.5	5.9
Widowed/surviving partner	42	8.4	Widowed	45	8.8	87	8.6	7.6
Refused/don't know			Refused/don't know	2	0.4	2	0.2	–
<b>Total</b>	<b>500</b>			<b>513</b>		<b>1013</b>	100	100

[Population: 2001 Census, Key Statistics for Health Areas, table KS03 (England)].

TABLE 46 Employment status

Status	OSCA		MOPSU		Total		Population (%)
	Frequency	%	Frequency	%	Frequency	%	
Working full time ( $\geq 30$ hours a week)	176	35.2	206	40.2	382	37.7	49.1
Working part time (<30 hours a week)	65	13	65	12.7	130	12.8	11.8
Full-time student	19	3.8	16	3.1	35	3.5	7.3
Part-time student	4	0.8	5	1.0	9	0.9	
Not working – looking for work	28	5.6	23	4.5	51	5.0	3.4
Not working – not looking for work	10	2	10	1.9	20	2.0	
Not working – unable for medical reasons	31	6.2	23	4.5	54	5.3	5.3
Retired	120	24	125	24.4	245	24.2	13.5
Looking after home	35	7	37	7.2	72	7.1	6.5
Other	12	2.4	3	0.6	15	1.5	3.1
<b>Total</b>	<b>500</b>	<b>100</b>	<b>513</b>	<b>100</b>	<b>1013</b>	<b>100</b>	<b>100</b>

[Population: 2001 Census, Key Statistics for Health Areas, table KS09 (England)].

TABLE 47 Ethnicity

Ethnicity	OSCA		MOPSU		Total		Population (%)
	Frequency	%	Frequency	%	Frequency	%	
White	422	84.4	419	81.7	841	83.0	90.9
Asian	39	7.8	51	9.9	90	8.9	5.0
Black	24	4.8	32	6.2	56	5.5	2.3
Any other background	14	2.8	11	2.1	25	2.5	1.8
Declined to answer	1	0.2			1	0.1	–
<b>Total</b>	<b>500</b>	<b>100</b>	<b>513</b>	<b>100</b>	<b>1013</b>	<b>100</b>	<b>100</b>

[Population: 2001 Census, Key Statistics for Health Areas, table KS06 (England)].

the sample appears to be broadly representative; this would suggest that we have a significantly higher proportion of people that have taken early retirement than would have otherwise been expected.

The ethnic profile of the sample (*Table 47*) is encouraging, and in fact under-represents white individuals and over-represents Asian and black individuals. This is useful, as it provides more data to power an analysis of the impact of ethnicity on preferences.

*Table 48* shows the number of children under the age of 16 years who live in the household. Reflecting other characteristics of the samples, there is under-representation of households with children.

*Table 49* shows the annual household income distribution of respondents. It is noteworthy that the level of refusal was reduced significantly between the MOPSU and OSCA study, which is believed to be the result of a change in the interviewer briefing instructions. The income bands used between the studies are not aligned, so it is difficult to make comparisons; however, it would appear that the survey samples under-represent higher-income households, probably reflecting the employment status of our samples.

*Table 50* shows the composition of the samples by area of residence, although it should be noted that we would not expect the sample to be representative within these classifications, as it was recruited from four distinct geographic areas.

*Table 51* shows how respondents rated their health at the time of the interview. The health status looks to be broadly the same across the OSCA and MOPSU samples, with the combined sample having 73.2% of respondents who rated their health being good or better. Only four respondents claimed to have very bad health.

Finally, *Table 52* suggests that the sample under-represents those stating that they were receiving disabled person's tax credit, housing benefit and pension credit. It is not clear whether this under-representation is a genuine skew in the sample or whether there is an issue related to respondents not being willing to state which benefits they were claiming (which in some cases they might not be formally entitled to claim).

The samples in the OSCA and MOPSU surveys were very similar, and broadly consistent with the population in all of these observable dimensions.

## Time trade-off follow-up sample

In the following tables (*Tables 53–56*) we compare the composition of the subsample of the general population recruited to participate in the follow-up TTO survey with the sample that were interviewed in the main preference study. The tables present the proportion of the sample within each category.

From these tables, we can see that the TTO sample has a broadly similar composition in terms of gender, age, socioeconomic status and annual household income. The TTO sample contains a slightly lower proportion of male respondents; however, in all other respects the samples appear comparable. We therefore have no reason to suspect that the sample recruitment for the TTO survey, or the self-selection of those willing to participate in this follow-up survey, will have introduced any biases in the estimation of the SCRQoL preferences.

**TABLE 48** Number of children in the household

No. of children	OSCA		MOPSU		Total		Population (%)
	Frequency	%	Frequency	%	Frequency	%	
0	321	64.2	327	63.7	648	63.9	56.6
1	97	19.4	81	15.8	178	17.6	18.5
2	68	13.6	77	15.0	145	14.3	17.2
3	10	2.0	18	3.5	28	2.8	7.7
4	3	0.6	8	1.6	11	1.1	
≥5	1	0.2	2	0.4	3	0.3	
<b>Total</b>	<b>500</b>	<b>100</b>	<b>513</b>	<b>100</b>	<b>1013</b>	<b>100</b>	<b>100</b>

[Population: 2001 Census, National report for England and Wales, table S007 (England and Wales)].

**TABLE 49** Annual household income

Income (£)	OSCA		MOPSU		Total		Population (%)
	Frequency	%	Frequency	%	Frequency	%	
Under 6499	36	9.1	29	8.9	65	9.0	50.0 <sup>a</sup>
6500–9499	42	10.7	36	11.0	78	10.8	
9500–30,000	200	50.8	173	53.1	373	51.8	
30,001–50,000	78	19.8	61	18.7	139	19.3	31.0 <sup>b</sup>
50,001–75,000	31	7.9	16	4.9	47	6.5	19.0
>75,000	7	1.8	11	3.4	18	2.5	
<b>Stated income</b>	<b>394</b>		<b>326</b>		<b>720</b>		
Refused/don't know	50		76		126		
Refused	56		111		167		
<b>Total</b>	<b>500</b>	<b>100</b>	<b>513</b>	<b>100</b>	<b>1013</b>	<b>100</b>	<b>100</b>

a The actual range is up to £26,000.

b The actual range is £26,000–52,000.

[Population: 2007/08 Family Resource Survey, table 3.6 (England)].

**TABLE 50** Area of residence

Area	OSCA		MOPSU		Total		Population (%)
	Frequency	%	Frequency	%	Frequency	%	
A London borough	121	24.2	125	24.4	246	24.3	14.7
Another city/large town	96	19.2	75	14.6	171	16.9	66.3
Suburbs of a city/large town	130	26	82	16.0	212	20.9	
A small town	39	7.8	119	23.2	158	15.6	9.1
A rural area or village	114	22.8	112	21.8	226	22.3	9.8
<b>Total</b>	<b>500</b>	<b>100</b>	<b>513</b>	<b>100</b>	<b>1013</b>	<b>100</b>	<b>100</b>

[Population: 2001 Census, Key Statistics, table KS20 by rural and urban classification (England)].

**TABLE 51** Self-reported health status

Health status	OSCA		MOPSU		Total		Population (%)
	Frequency	%	Frequency	%	Frequency	%	
Very good	153	30.6	151	29.4	303	29.9	68.8
Good	212	42.4	227	44.2	439	43.3	
Fair	108	21.6	107	20.9	215	21.2	22.2
Bad	24	4.8	27	5.3	51	5.1	9.0
Very bad	3	0.6	1	0.2	4	0.4	
<b>Total</b>	<b>500</b>	<b>100</b>	<b>513</b>	<b>100</b>	<b>1013</b>	<b>100</b>	<b>100</b>

[Population: 2001 Census, Key Statistics for Health Areas, table KS08 (England)].

**TABLE 52** Respondent or household member who received benefits

Benefit	OSCA		MOPSU		Total		Population (%)
	Frequency	%	Frequency	%	Frequency	%	
Income support	35	7.0	39	7.6	74	7.3	16.0
Working families' tax credit	52	10.4	38	7.4	90	8.9	
Income-based Jobseeker's Allowance	21	4.2	16	3.1	37	3.7	3.0
Disabled person's tax credit	2	0.4	9	1.8	11	1.1	15.0
Housing benefit	72	14.4	62	12.1	134	13.2	20.0
Council tax benefit	76	15.2	81	15.8	157	15.5	16.0
Pension credit	21	4.2	31	6.0	52	5.1	29.0
Attendance allowance	12	2.4	–	–	–	–	–
Carer's allowance	25	5.0	–	–	–	–	–
Disability living allowance	44	8.8	–	–	–	–	–
Employment and support allowance	9	1.8	–	–	–	–	–
Incapacity benefit	20	4.0	–	–	–	–	–
Did not receive any benefit	305	61.0	362	70.6	667	65.8	32.0

[Population: 2000/01 Regional Trends 37, table KS8.8 (England)].

**TABLE 53** Gender distribution

Gender	BWS sample	TTO
Male	49.0	46.0
Female	51.0	54.0
<b>Total</b>	<b>100.0</b>	<b>100.0</b>

**TABLE 54** Age distribution

Age (years)	BWS sample	TTO
18–24	11.4	10.3
25–30	10.8	11.9
31–39	17.8	20.6
40–49	15.6	14.3
50–64	24.2	22.2
65–69	7.4	8.7
70–79	8.8	9.5
80–89	3.8	2.4
≥90	0.2	0
<b>Total</b>	<b>100.0</b>	<b>100.0</b>

**TABLE 55** Socioeconomic status

Band	BWS sample	TTO
A	1.0	1.6
B	24.6	25.4
C1	29.0	29.4
C2	19.2	26.2
D	11.8	7.1
E	14.4	10.3
<b>Total</b>	<b>100.0</b>	<b>100.0</b>

**TABLE 56** Annual household income

Income (£)	BWS sample	TTO
<6499	9.1	4.0
6500–9499	10.7	10.0
9500–30,000	50.8	53.0
30,001–50,000	19.8	25.0
50,001–75,000	7.9	8.0
>75,000	1.8	0.0
<b>Total (with stated income)</b>	<b>100.0</b>	<b>100.0</b>

## The service user sample

Services users were recruited through the 2010 UES, which was conducted among people who had received equipment services between September and December 2009. Nine councils took part from a variety of regions and LA types, sending information about the OSCA study to respondents of that survey. The fieldwork organisation, Accent, was sent contact information about people who expressed an interest in participating. Some problems were encountered in following up people, which is not unexpected given the nature of the service user population. A total of 1032 contacts were attempted from a sample of 1106 respondents who were willing to participate in further research, which produced 458 (44%) complete interviews. Non-respondents comprised those who were not contactable ( $n = 500$ , 48%), refusals ( $n = 58$ , 6%) and those who had died ( $n = 16$ , 2%).

Key demographic and socioeconomic characteristics are shown in *Tables 57–67*. Where comparable data are available, we identify these for the national survey, which had 65,000 respondents.<sup>13</sup>

From *Table 57*, we can see that the majority of service users in the sample were female, very close to the national picture (35.2%).<sup>13</sup>

*Table 58* shows the majority of individuals in the sample were aged  $\geq 50$  years. Not surprisingly, more than half of the sample consists of respondents between 70 and 89 years of age. The sample was slightly older than respondents nationally: just under one-fifth (18.8%) were aged  $< 65$  years compared with 20.9% nationally.<sup>13</sup>

*Table 59* shows that, as we would expect given the age distribution and means-tested nature of social care, a much higher proportion had no formal qualification compared with the general population: nearly two-thirds of the sample, compared with less than one-fifth nationally.

As shown in *Table 60*, there was only one respondent out of 458 in the highest SEG. Approximately half of the sample belonged to band C.

*Table 61* shows that over two-fifths of service users were married (or equivalent) and a similar proportion were widowed or surviving partners.

*Table 62* shows that, as expected, the vast majority of service users were retired or unable to work for medical reasons.

*Table 63* shows that nearly all service users in the sample were white: higher than the national picture. This would be expected as a result of the older age profile of people using equipment, but it was also a slightly higher proportion than respondents to the national survey, where 93.5% were white.<sup>13</sup>

*Table 64* shows the annual income distribution of service users in the sample. More than 50% of respondents had an annual household income of  $< \pounds 14,400$ . The level of refusal was comparable with the general population in this study.

*Table 65* shows the composition of the samples by area of residence. A higher proportion of respondents were living either in a small town or a rural area or village compared with the national picture.

*Table 66* shows that, as we might expect, the sample were much less likely to report themselves in good health than the general population, with the majority of respondents rating themselves as having 'fair' or worse health.

Finally, *Table 67* shows, again as we might expect, a higher proportion of service users were likely to be receiving welfare benefits than the general population.

**TABLE 57** Gender distribution

Gender	OSCA		Population (%)
	Frequency	%	
Male	160	34.9	48.6
Female	298	65.1	51.3
<b>Total</b>	<b>458</b>	<b>100</b>	<b>100</b>

[Population: Office for National Statistics mid-2008 estimate, table 4 (England)].

**TABLE 58** Age distribution

Age (years)	OSCA		Population (%)
	Frequency	%	
18–24	1	0.2	12.1
25–30	0	0.0	10.1
31–39	3	0.7	15.8
40–49	13	2.8	18.8
50–64	69	15.1	22.8
65–69	49	10.7	5.6
70–79	124	27.1	9.0
80–89	168	36.7	4.9
≥ 90	31	6.8	0.9
<b>Total</b>	<b>458</b>	<b>100</b>	<b>100</b>

[Population: Office for National Statistics mid-2008 estimate, table 4 (England)].

**TABLE 59** Level of education

Level of education	OSCA		Population (%)
	Frequency	%	
No formal qualification	282	61.6	18.9
GCSE/O-level	48	10.5	34.4
A-levels or equivalent	19	4.1	12.4
Professional qualification below degree level	44	9.6	8.7
Degree level qualification or equivalent	23	5.0	7.8
Higher degree	11	2.4	3.5
Other	31	6.8	14.4
<b>Total</b>	<b>458</b>	<b>100</b>	<b>100</b>

[Population: 2001 Census, National report for England and Wales, table S115 (England and Wales)].

**TABLE 60** Socioeconomic status

Band	OSCA		Population (%)
	Frequency	%	
A	1	0.2	22.0
B	79	17.2	
C1	129	28.2	29.7
C2	97	21.2	15.1
D	48	10.5	17.2
E	100	21.8	16.1
Not stated	4	0.9	–
<b>Total</b>	<b>458</b>	<b>100</b>	<b>100</b>

[Population: 2001 Census, National report for England and Wales, table S066 (England and Wales)].

**TABLE 61** Marital status

Status	OSCA		Population (%)
	Frequency	%	
Married, or equivalent	196	42.8	61.5
Never been married, or equivalent	25	5.5	23.1
Separated (from spouse or partner)	6	1.3	1.9
Divorced	37	8.1	5.9
Widowed/surviving partner	192	41.9	7.6
Refused/don't know	2	0.4	–
<b>Total</b>	<b>458</b>	<b>100</b>	<b>100</b>

[Population: 2001 Census, Key Statistics for Health Areas, table KS03 (England)].

**TABLE 62** Employment status

Status	OSCA		Population (%)
	Frequency	%	
Working full time ( $\geq 30$ hours a week)	3	0.7	49.1
Working part-time ( $< 30$ hours a week)	7	1.5	11.8
Full-time student	0	0.0	7.3
Part-time student	1	0.2	
Not working – looking for work	0	0.0	3.4
Not working – not looking for work	1	0.2	
Not working – unable for medical reasons	54	11.8	5.3
Retired	389	84.9	13.5
Looking after home	2	0.4	6.5
Other	1	0.2	3.1
<b>Total</b>	<b>458</b>	<b>100</b>	<b>100</b>

[Population: 2001 Census, Key Statistics for Health Areas, table KS09 (England)].



**TABLE 63** Ethnicity

Ethnicity	OSCA		Population (%)
	Frequency	%	
White	452	98.7	90.9
Asian	5	1.1	5.0
Black	1	0.2	2.3
Any other background	0	0.0	1.8
Declined to answer	0	0.0	-
<b>Total</b>	<b>458</b>	<b>100</b>	<b>100</b>

[Population: 2001 Census, Key Statistics for Health Areas, table KS06 (England)].

**TABLE 64** Annual household income

Income (£)	OSCA		Population (%)
	Frequency	%	
<6499	34	7.4	28 <sup>a</sup>
6500–9499	105	22.9	
9500–14,400	121	26.4	
14,401–19,500	38	8.3	22 <sup>b</sup>
19,501–22,000	23	5.0	
22,001–30,000	18	3.9	
30,001–50,000	9	2.0	31.0 <sup>c</sup>
50,001–75,000	1	0.2	19.0
>75,000	0	0	
Don't know	51	11.1	
Refused	58	12.7	
<b>Total</b>	<b>458</b>	<b>100</b>	<b>100</b>

a The actual range is up to £15,600.

b The actual range is £15,600 to <£26,000.

c The actual range is £26,000–52,000.

[Population: 2007/08 Family Resource Survey, table 3.6 (England)].

**TABLE 65** Area of residence

Type of area	OSCA		Population (%)
	Frequency	%	
A London borough	10	2.2	14.7
Another city/large town	77	16.8	66.3
Suburbs of a city/large town	42	9.2	
A small town	172	37.6	9.1
A rural area or village	157	34.3	9.8
<b>Total</b>	<b>458</b>	<b>100</b>	<b>100</b>

[Population: 2001 Census, Key Statistics, table KS20 by rural and urban classification (England)].

**TABLE 66** Self-reported health status

Status	OSCA		Population (%)
	Frequency	%	
Very good	18	3.9	68.8
Good	107	23.4	
Fair	200	43.7	22.2
Bad	111	24.2	9.0
Very bad	22	4.8	
<b>Total</b>	<b>458</b>	<b>100</b>	<b>100</b>

[Population: 2001 Census, Key Statistics for Health Areas, table KS08 (England)].

**TABLE 67** Respondent or household member who received benefits

Benefit	OSCA		Population (%)
	Frequency	%	
Income support	22	4.8	16.0
Working families' tax credit	5	1.1	
Income-based Jobseeker's Allowance	1	0.2	3.0
Disabled person's tax credit	15	3.3	15.0
Housing benefit	125	27.3	20.0
Council tax benefit	193	42.1	16.0
Pension credit	124	27.1	29.0
Attendance allowance	178	38.9	-
Carer's allowance	91	19.9	-
Disability living allowance	169	36.9	-
Employment and support allowance	2	0.4	-
Incapacity benefit	44	9.6	-
Did not receive any benefit	70	15.3	32.0

[Population: 2000/01 Regional Trends 37, table KS8.8 (England)].

## Service user follow-up sample

**TABLE 68** Sociodemographic and economic characteristics of follow-up sample of 101 equipment service users

Characteristic	Frequency	%
<b>Sex (n = 101)</b>		
Female	61	60
<b>Ethnicity (n = 101)</b>		
White	101	100
<b>Age, years (n = 101)</b>		
18–24	1	1
40–49	2	2
50–64	18	18
65–69	12	12
70–79	29	29
80–89	35	35
≥ 90	4	4
<b>Area of residence (n = 101)</b>		
London borough	0	0
Another large city or town	21	21
Suburbs of large city/town	6	6
Small town	38	38
Rural area or village	36	36
<b>Marital status (n = 101)</b>		
Married/living together	53	52
Never married	6	6
Widowed	34	34
Separated/divorced	8	8
<b>Living situation (n = 101)</b>		
Live alone		
<b>Tenure (n = 100)</b>		
Owner-occupier	70	70
Rent-paying tenant	28	28
Tenant living rent free	2	2
<b>Income, per week (£) (n = 79)</b>		
≤ 275	61	77
276–374	7	9
375–424	4	5
425–574	4	5
≥ 575	3	4



## Appendix 7

# Domain-level weights of a collapsed three-option version of Adult Social Care Outcome Toolkit

In order to provide a flexible approach to measuring social care outcomes, ASCOT has a number of versions of the outcome measure ([www.pssru.ac.uk/ASCOT](http://www.pssru.ac.uk/ASCOT)). One of these uses a multimethod approach to establishing SCRQoL states for people living in care homes, who often have a limited capacity to respond to interviews or questionnaires. This version has only three levels in each domain, reflecting no, some or low-level needs and high needs. In order to allow us to estimate preference weights for this version, we investigated collapsing the four-level measure into a three-level version by combining the two top two 'no needs' levels. The following sections presented these models using the general population and service users' data, respectively.

### General population

*Table 69* presents the general population model in which the top two levels of each domain were combined. As shown in *Figure 25*, the estimated domain weights fell within the original values for the separate levels.

### Service users

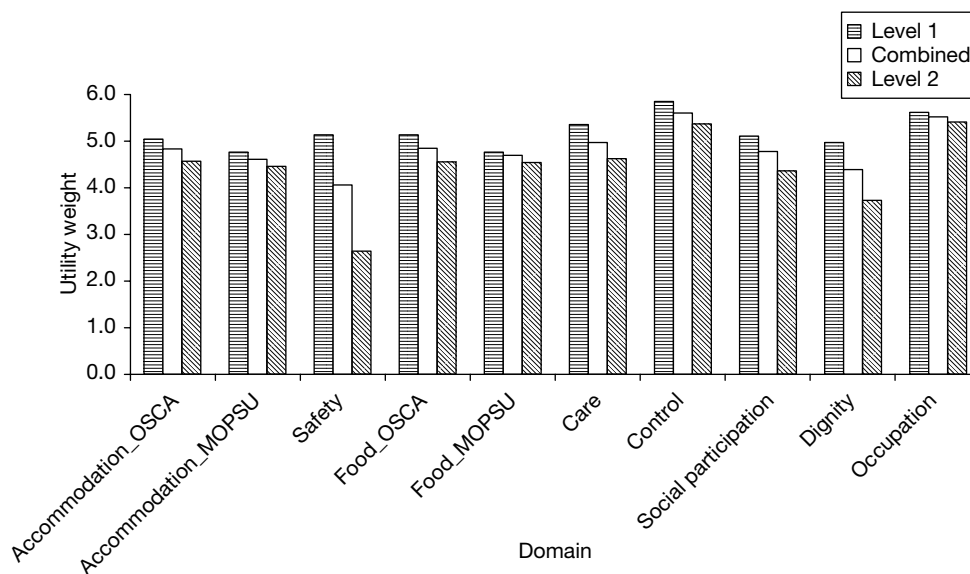
*Table 70* presents the service users' model in which the top two levels of each domain were combined. *Figure 26* shows that when the four-level measure is collapsed to a three-level measure, the estimated weights fall between the original two values of the separate levels.

**TABLE 69** General population domain level weights when four-level domains collapse to three-level domains

Domain	Coefficient	t-ratio (robust)
<b><i>Accommodation cleanliness and comfort</i></b>		
1. My home is as clean and comfortable as I want	4.81	27.3
2. My home is adequately clean and comfortable		
3. My home is not quite clean or comfortable enough	2.25	16.8
4. My home is not at all clean or comfortable	1.75	13.9
1. My home is as clean and comfortable as I want (MOPSU)	4.61	26.5
2. My home is adequately clean and comfortable (MOPSU)		
3. My home is less than adequately clean or comfortable (MOPSU)	2.35	17.1
4. My home is not at all clean or comfortable (MOPSU)	1.80	14.9
<b><i>Safety</i></b>		
1. I feel as safe as I want	4.04	24.0
2. Generally I feel adequately safe, but not as safe as I would like		
3. I feel less than adequately safe	1.79	16.4
4. I don't feel at all safe	0.68	7.5
<b><i>Food and drink</i></b>		
1. I get all the food and drink I like when I want	4.86	26.9
2. I get adequate food and drink at OK times		
3. I don't always get adequate or timely food and drink	1.80	14.6
4. I don't always get adequate or timely food and drink, and I think there is a risk to my health	1.13	9.7
1. I get all the food and drink I like when I want (MOPSU)	4.69	25.7
2. I get food and drink adequate for my needs (MOPSU)		
3. I don't get all the food and drink I need, but I don't think there is a risk to my health (MOPSU)	2.16	16.7
4. I don't get all the food and drink I need, and I think there is a risk to my health (MOPSU)	1.31	11.4
<b><i>Personal care</i></b>		
1. I feel clean and am able to present myself the way I like	5.00	26.7
2. I feel adequately clean and presentable		
3. I feel less than adequately clean or presentable	1.58	15.0
4. I don't feel at all clean or presentable	1.19	12.5
<b><i>Control over daily life</i></b>		
1. I have as much control over my daily life as I want	5.61	27.3
2. I have adequate control over my daily life		
3. I have some control over my daily life but not enough	3.21	21.3
4. I have no control over my daily life	0.00	Base level
<b><i>Social participation and involvement</i></b>		
1. I have as much social contact as I want with people I like	4.79	26.9
2. I have adequate social contact with people		
3. I have some social contact with people, but not enough	2.97	20.4
4. I have little social contact with people and feel socially isolated	1.48	14.6
<b><i>Dignity</i></b>		
1. The way I'm helped and treated makes me think and feel better about myself	4.40	24.5
2. The way I'm helped and treated does not affect the way I think or feel about myself		
3. The way I'm helped and treated sometimes undermines the way I think and feel about myself	1.81	16.5
4. The way I'm helped and treated completely undermines the way I think and feel about myself	1.60	15.7

**TABLE 69** General population domain level weights when four-level domains collapse to three-level domains (continued)

Domain	Coefficient	t-ratio (robust)
<b>Occupation and employment</b>		
1. I'm able to spend my time as I want, doing things I value or enjoy	5.54	26.8
2. I'm able do enough of the things I value or enjoy with my time		
3. I do some of the things I value or enjoy with my time but not enough	3.45	22.0
4. I don't do anything I value or enjoy with my time	1.05	12.0
<b>Domain position in the BWS task</b>		
Top: Respondent chose the domain that appeared first	0.19	4.6
Top two: Respondent chose the domain that appeared second	0.09	2.5
Bottom: Respondent chose the attribute that appeared last	0.01	0.3
Bottom two: Respondent chose the attribute that appeared before the last	-0.04	-1.3
<b>Scale parameters</b>		
$\mu_{\text{excessive}}$ : MOPSU data with 'excessive randomisation'	0.83	3.5
$\mu_{\text{worst}}$ : data relates to a 'worst' or 'second-worst' choice	0.82	8.8
$\mu_{\text{not-imagine}}$ : data relates to individuals who could not put themselves in the imaginary positions described in the choices	0.80	3.1
$\mu_{\text{London and South-East}}$ : data relates to individuals living in London and South East	0.81	6.0
$\mu_{\text{bhealth}}$ : data relates to individuals with self-reported fair, bad or very bad health	0.88	3.0
$\mu_{\text{int. duration}<15\text{min}}$ : data relates to individuals' interview duration shorter than 15 minutes	0.87	2.8
<b>Model diagnostics</b>		
No. of observations	29,728	29,728
df	39	39
Final log-likelihood	-43,205.2	-43,205.2
Rho <sup>2</sup> (0)	0.217	0.217

**FIGURE 25** Weights of the top two domain levels in OSCA and the collapsed model (general population).

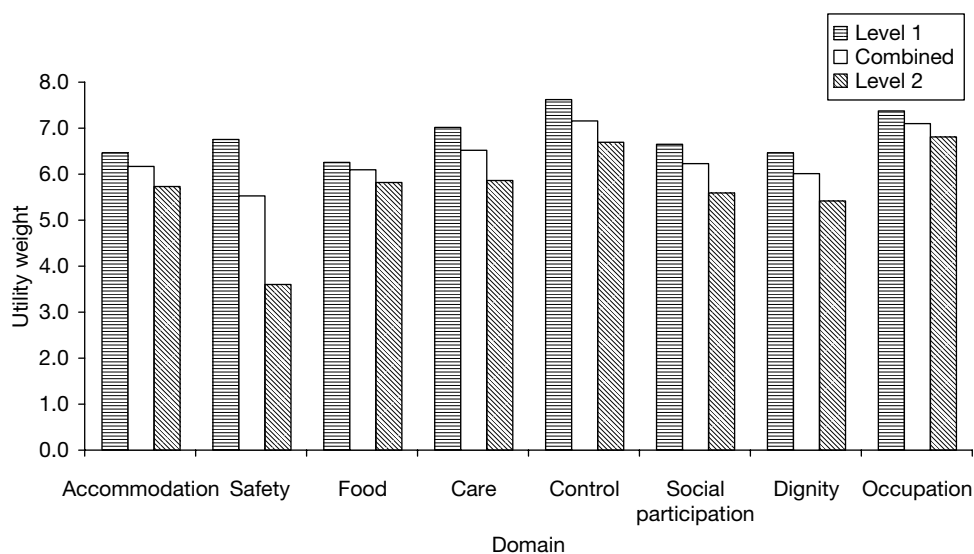
**TABLE 70** Service-user domain level weights when four-level domains collapse to three-level domains

Domain	Coefficient	t-ratio (robust)
<b><i>Accommodation cleanliness and comfort</i></b>		
1. My home is as clean and comfortable as I want	6.17	13.4
2. My home is adequately clean and comfortable		
3. My home is not quite clean or comfortable enough	3.30	9.9
4. My home is not at all clean or comfortable	2.74	9.1
<b><i>Safety</i></b>		
1. I feel as safe as I want	5.54	12.3
2. Generally I feel adequately safe, but not as safe as I would like		
3. I feel less than adequately safe	2.37	8.8
4. I don't feel at all safe	1.04	5.0
<b><i>Food and drink</i></b>		
1. I get all the food and drink I like when I want	6.11	12.9
2. I get adequate food and drink at OK times		
3. I don't always get adequate or timely food and drink	2.92	10.1
4. I don't always get adequate or timely food and drink, and I think there is a risk to my health	1.73	7.9
<b><i>Personal care</i></b>		
1. I feel clean and am able to present myself the way I like	6.52	13.4
2. I feel adequately clean and presentable		
3. I feel less than adequately clean or presentable	2.37	8.5
4. I don't feel at all clean or presentable	1.83	7.1
<b><i>Control over daily life</i></b>		
1. I have as much control over my daily life as I want	7.20	13.1
2. I have adequate control over my daily life		
3. I have some control over my daily life but not enough	4.56	11.1
4. I have no control over my daily life	0.00	Base level
<b><i>Social participation and involvement</i></b>		
1. I have as much social contact as I want with people I like	6.24	12.8
2. I have adequate social contact with people		
3. I have some social contact with people, but not enough	4.27	10.9
4. I have little social contact with people and feel socially isolated	2.22	8.4
<b><i>Dignity</i></b>		
1. The way I'm helped and treated makes me think and feel better about myself	6.02	12.5
2. The way I'm helped and treated does not affect the way I think or feel about myself		
3. The way I'm helped and treated sometimes undermines the way I think and feel about myself	2.86	10.0
4. The way I'm helped and treated completely undermines the way I think and feel about myself	2.43	9.2
<b><i>Occupation and employment</i></b>		
1. I'm able to spend my time as I want, doing things I value or enjoy	7.13	13.3
2. I'm able to do enough of the things I value or enjoy with my time		
3. I do some of the things I value or enjoy with my time but not enough	5.04	11.5
4. I don't do anything I value or enjoy with my time	1.74	7.5



**TABLE 70** Service-user domain level weights when four-level domains collapse to three-level domains (*continued*)

Domain	Coefficient	t-ratio (robust)
<b>Domain position in the BWS task</b>		
Top: respondent chose the domain that appeared first	0.38	5.2
Top two: respondent chose the domain that appeared second	0.18	2.9
Bottom: respondent chose the attribute that appeared last	0.04	0.6
Bottom two: respondent chose the attribute that appeared before the last	-0.06	-0.8
<b>Scale parameters</b>		
$\mu_{\text{worst}}$ : data relates to a 'worst' or 'second-worst' choice	0.63	12.6
$\mu_{\text{notSE}}$ : data relates to residents outside of the south-east of England	0.72	6.3
$\mu_{\text{O-level}}$ : data relates to individuals with no formal qualification or GCSE/O-level	0.82	3.2
<b>Model diagnostics</b>		
No. of observations	11,744	
df	30	
Final log-likelihood	-17,306.4	
Rho <sup>2</sup> (0)	0.206	

**FIGURE 26** Weights of the top two domain levels in OSCA and the collapsed model (service users).



## Appendix 8

# Choice elicitation and model estimation procedure in best–worst scaling data

### Specification of the model structure

Each best–worst task involved four sequential choices (i.e. best, worst, second best, second worst). First, respondents were presented with eight domain levels and were asked to choose what they thought would be the best. Next, the chosen domain level was taken off the best–worst task and respondents were asked to choose the worst of the remaining alternatives. Following that, respondents were again asked to choose the best (what we call second best) of the remaining alternatives (six in the OSCA experiment, respectively). Finally, respondents chose the worst alternative (what we call second worst) of the domain levels remaining in the given best–worst task.

There are two ways of analysing this: either by looking at it as respondents aiming to select those best–worst pairs that maximise the difference in utility (in two rounds of best–worst choices) (and set up equations where the alternatives are all possible combinations of best–worst–second best–second worst sequences), or starting from the point of considering these as four separate choice processes and then take account of the fact they come from the same original BWS scenario.

The Maxdiff model is the first approach described above.<sup>135</sup> However, the estimation procedure for the Maxdiff is significantly more complicated if used to analyse a data set that includes second-best and second-worst choices. Specifically, the problem lies around the very higher number of possible alternatives. In this study, the number of alternatives should include all possible combinations of best–worst pairs, namely  $2(8 \times 4)$ . When we add the second-best and second-worst choices, the number of alternatives quickly becomes impossible to handle – even for a simple estimation. Also, the Maxdiff approach is used less by researchers, and the majority of applications model BWS data as a series of sequential choices.<sup>134</sup>

Therefore, the analysis started from the assumption that all choices made were made sequentially, from successively smaller choice sets.<sup>96,103,134,135</sup> The model coding involves the following:

- *choice of best* respondent chooses one out of eight alternatives
- *choice of worst* respondent chooses one out of seven alternatives
- *choice of second best* respondent chooses one out of six alternatives
- *choice of second worst* respondent chooses one out of five alternatives.

In the first case, the respondent is indicating the alternative with the highest utility, in the second the alternative with the lowest utility, then next highest utility, then next lowest utility.

In principle, the alternatives in the model are every possible domain level, but then with availability criteria that say that in a given choice scenario a respondent sees only eight (or seven or six or five) of the possible domain levels. It is possible to further simplify the coding (and gain significant efficiencies in estimating the model), as only one level appears for any given domain

at the same time, so it is possible to set up a model with eight possible alternatives to choose between, which have utility functions defined to take account of which level is presented for that domain. This is done using dummy coding to reflect the domain levels that were presented.

The terms in the utility are positive when the outcome is a best or second-best choice (i.e. choice is maximising positive utility) and negative where the outcome is a worst or second-worst choice (i.e. choice is maximising negative utility).

Additional availability criteria are applied moving through the best, worst, second-best, second-worst sequence to eliminate the previously chosen domains from the choice set for the subsequent choices.

Therefore, the dependent variable in the model is the domain level chosen (from those presented) and the utility of the alternative is a function of the level at which the domain was presented. The coefficient on just one of all domain levels needs to be constrained to zero to allow the model to be identified, and all other coefficients are therefore relative to this domain level.

### Extension of this model to explain differences in scale heterogeneity

One of the main concepts in the field of random utility modelling is the notion of scale of the utility. In particular, and looking at the case of a MNL model, the probability of choosing alternative  $i$  out of  $j$  alternatives is given by:

$$P(i) = \exp(\mu \times Vi) / [\sum_j \exp(\mu \times Vj)] \quad [\text{Equation 8}]$$

where:  $V_i$  gives the modelled utility of alternative  $i$  (e.g.  $\beta'xi$ ), and  $\mu$  is the scale, which is inversely proportional to the variance of the error terms, where the variance of the Gumbel distribution is given by  $\pi^2/(6 \times \mu^2)$ .<sup>111</sup>

As the scale increases (variance decreases), respondents become more sensitive to the explanatory variables and the choice process becomes more deterministic. In practice, an increase in scale equates to higher marginal utility coefficients. It is not possible to identify the scale separately from the coefficients, and therefore in estimation the scale is arbitrarily set to 1. So, an increase in scale can equally well be achieved by an increase in  $\mu$  or in  $\beta$ . However, when dealing with multiple data sources, or multiple types of responses, it is possible to estimate separate scales for different segments of the estimation data.<sup>105,136</sup>

In the present context, on the basis of past experience we expect that the scale is highest (lowest error variance) for the first choice (i.e. best level) and is lower for the remaining three decisions in each BWS exercise. Early results showed this to be true, but with little or no difference between the scale for best and second-best choice, and between the worst and second-worst choice, meaning that we can set  $\mu_{\text{best}}$  to 1 (required for identification reasons) and estimate  $\mu_{\text{worst}}$ , where the former is associated with the best and second-best choices, and the latter with the worst and second-worst choices.

Results from the MOPSU study also showed that the model scale for those respondents who received the excessively randomised version of the survey was lower (higher variance in the error term), and therefore we incorporated a separate scale,  $\mu_{\text{excessive}}$ , for those respondents who received the version of the survey with excessive randomisation.

The question then arises as to how to deal with multiple scale differences, which can occur jointly. A possibility would be to use four separate scale parameters, namely  $\mu_{\text{best-standard}}$  (set to 1),  $\mu_{\text{best-excessive}}$ ,  $\mu_{\text{worst-standard}}$ , and  $\mu_{\text{worst-excessive}}$ . This specification allows for a differential scale impact of the two types of randomisation on best and worst choices, and early results in the MOPSU study showed that this degree of flexibility was not required. Instead, we estimated two scale effects: worst (relative to best) and excessive (relative to standard), and applied these jointly as necessary. Here, with the expectation that the scale for worst and second-worst choices is lower than for best and second-best choices, and that the scale for observations from the excessively randomised survey is similarly lower, we would have that  $\mu_{\text{worst}} < 1$  and  $\mu_{\text{excessive}} < 1$ , which would mean even lower scale for worst and second-worst choices in the excessively randomised survey. This approach was carried forwards to the main OSCA analysis, where many more scales were considered simultaneously.

## Dealing with the repeated choice nature of the data

In the most basic specification, the repeated choice nature of the data is not taken into account, and each observation is treated independently. This assumption is clearly incorrect, and three possible cases of correlation arise.

This issue was considered in some detail in the MOPSU study. In possibly the most obvious extension, a model can be specified that recognises the fact that the set of 32 (four choices  $\times$  eight choice scenarios) for a given individual are in fact correlated. An extension of this comes when we additionally recognise the special relationship between the four choices made during each BWS exercise, meaning that the choices for a given individual are explicitly recognised as a set of eight sequences of four choices. A third approach, which is a simplification of the second case, which was useful for testing purposes as detailed later on, ignores the relationship between the eight sets of BWS exercises, and only accounts for the relationship between the four choices within each best–worst task.

At present, there is no clear guidance among choice modellers on how the repeated choice nature of panel data should be taken into account. A number of applications make use of a post-estimation correction approach, generally leading to an upwards correction in the SEs. However, these methods, especially resampling approaches such as Jackknife or Bootstrap, are somewhat arbitrary, and sensitive to assumptions made during specification. Additionally, in the case where individual model runs already take several hours, as is the case here owing to the multiplicative scale approach, resampling approaches can be computationally very expensive. Recently, there has been a trend to make use of random coefficients models, primarily Mixed Logit, where the panel nature of the data is taken into account by allowing for variation in coefficients across respondents while keeping the coefficients constant across choices for the same respondent. In many studies, this one included, taste heterogeneity is not the main topic of investigation, however, and, indeed, analysts may only be interested in point estimates for average values of the coefficients, which causes complications when relying on random coefficients models. Several studies have attempted to avoid this issue by relying on an error components approach to accommodate correlation across choices without introducing a representation of taste heterogeneity, but issues with specification and identification arise, alongside a very significant rise in estimation cost.

In the present context, an error components approach would be possible, with integration being carried out at two different levels, namely at the level of an individual and at the level of individual BWS exercises, i.e. having:

$$L = \prod_n \prod_s \prod_t P_{nst} \quad [\text{Equation 9}]$$

where  $n$  is the index over respondents,  $s$  is the index of BWS exercises, and  $t$  is the index for individual choices within an exercise.

The fact that integration is carried out in two places allows the two types of correlation to be accommodated, but also leads to an almost insurmountable rise in estimation cost, especially when taking into account the fact that a very large number of error components (and hence dimensions of the integral) would be required given the high number of alternatives (individual attribute-level combinations). We therefore reject the error-component approach in this context.

The solution put forward here is to limit ourselves to the outer panel, i.e. treating the 32 choices for one respondent as a block, but without special treatment for the four choices within a given BWS experiment. This assumption is, however, only justified if we can first establish that this outer panel is the main (and ideally only) reason for correlation across choices.

We make the common assumption that the estimates from the naive model are consistent and that only the SEs need to be corrected. The correction approach used to accommodate the impact of the panel nature of the data on the SEs is the sandwich estimator,<sup>135</sup> which allows for a correction of the SEs owing to some kinds of misspecification. In particular, the sandwich estimator is defined as:

$$S = (-H) - 1 \times B \times (-H) - 1 \quad [\text{Equation 10}]$$

where  $H$  is the Hessian, i.e. the matrix of second derivatives, and  $B$  is the  $BHHH$  matrix, which is given by the covariance matrix of the first derivatives, calculated over the observations on which the likelihood is calculated.

This differs from the classical covariance matrix, which is given by  $(-H) - 1$ , in the case of misspecification, i.e. when  $-H \neq B$ . If the model is specified correctly, the two approaches give the same covariance matrix. As Train<sup>135</sup> states, ‘this [covariance] matrix obtained is called “the robust covariance matrix” since it is valid whether or not the model is correctly specified’.

Misspecification can obviously be caused by a multitude of factors, linked to the specific assumptions made in model specification. As a first step, a base model was thus estimated with no recognition of the panel nature of the data, and the classical covariance matrix was compared to the robust covariance matrix. Within the MOPUSU study we found that the SEs with the robust specification were slightly higher but, on average, only by around 6%, indicating some minor misspecification.

As mentioned above, the specific interest in using the sandwich estimator in the present study was to address the potential downwards bias in the SEs caused by not accounting for the repeated choice nature of the data. We now return to the above statement for  $S$ , which is a function of  $H$  and  $B$ . When estimating a model with panel data, it is possible to base  $B$  on sequences of choices instead of individual choices, leading to different results for  $S$ , where this is not the case for the classical covariance matrix which is identical whether individual choices or sequences of choices are used.

Three different models were estimated in the exploratory work in the MOPUSU study. In the first model, we made use of individual choice probabilities in the calculation of  $S$ . This was followed by a model in which we made use of the full set of 32 choices for each respondent, which was observed to lead to a noticeable (further) upwards correction of the SEs, by on average of 27%

across all parameters (comparing the robust SEs for the base model with those for the model taking into account the panel nature of the data). In contrast, when working with sets of four choices – i.e. looking at individual BWS exercises – we observed an upwards correction of only 3%. This suggests very little or no effects for the small panel, especially when considering that recognising only the small panel potentially captures some of the effects of the large panel (by still grouping together some of the choices for a given respondent), meaning that the effects for the small panel are possibly even smaller than observed. On this basis, the assumption to focus solely on the large panel seems justified, while conceding that a further very small upwards correction in the SEs would be obtained by accommodating the small panel.

The preliminary MNL models, with scaling in only single dimensions, were estimated and developed using Alogit 4.3<sup>137</sup> (HCG Software, London) and the scaled MNL models with the robust SEs estimated using the sandwich estimator (reported throughout this report) were estimated using Biogeme 1.8.<sup>138</sup>

To summarise, the approach adopted within this study to deal with the repeated choices nature of the data has been to specifying the model such that the *BHHH* matrix can also be estimated on the sequence of choices from each individual. This allows the use of the sandwich estimator, which provides more robust estimates of the SEs, taking in to account the panel nature of the data.





# Health Technology Assessment programme

**Director,**  
**Professor Tom Walley, CBE,**  
 Director, NIHR HTA programme,  
 Professor of Clinical Pharmacology,  
 University of Liverpool

**Deputy Director,**  
**Professor Hywel Williams,**  
 Professor of Dermato-Epidemiology,  
 Centre of Evidence-Based Dermatology,  
 University of Nottingham

## Prioritisation Group

### Members

<p><b>Chair,</b>  <b>Professor Tom Walley, CBE,</b>          Director, NIHR HTA programme, Professor of Clinical Pharmacology, University of Liverpool</p> <p>Professor Imti Choonara,          Professor in Child Health, Academic Division of Child Health, University of Nottingham          Chair – Pharmaceuticals Panel</p> <p>Dr Bob Coates,          Consultant Advisor – Disease Prevention Panel</p> <p>Dr Andrew Cook,          Consultant Advisor – Intervention Procedures Panel</p> <p>Dr Peter Davidson,          Director of NETSCC, Health Technology Assessment</p>	<p>Dr Nick Hicks,          Consultant Adviser – Diagnostic Technologies and Screening Panel,          Consultant Advisor–Psychological and Community Therapies Panel</p> <p>Ms Susan Hird,          Consultant Advisor, External Devices and Physical Therapies Panel</p> <p>Professor Sallie Lamb,          Director, Warwick Clinical Trials Unit, Warwick Medical School, University of Warwick          Chair – HTA Clinical Evaluation and Trials Board</p> <p>Professor Jonathan Michaels,          Professor of Vascular Surgery, Sheffield Vascular Institute, University of Sheffield          Chair – Interventional Procedures Panel</p>	<p>Professor Ruairidh Milne,          Director – External Relations</p> <p>Dr John Pounsford,          Consultant Physician, Directorate of Medical Services, North Bristol NHS Trust          Chair – External Devices and Physical Therapies Panel</p> <p>Dr Vaughan Thomas,          Consultant Advisor – Pharmaceuticals Panel, Clinical Lead – Clinical Evaluation Trials Prioritisation Group</p> <p>Professor Margaret Thorogood,          Professor of Epidemiology, Health Sciences Research Institute, University of Warwick          Chair – Disease Prevention Panel</p>	<p>Professor Lindsay Turnbull,          Professor of Radiology, Centre for the MR Investigations, University of Hull          Chair – Diagnostic Technologies and Screening Panel</p> <p>Professor Scott Weich,          Professor of Psychiatry, Health Sciences Research Institute, University of Warwick          Chair – Psychological and Community Therapies Panel</p> <p>Professor Hywel Williams,          Director of Nottingham Clinical Trials Unit, Centre of Evidence-Based Dermatology, University of Nottingham          Chair – HTA Commissioning Board          Deputy HTA Programme Director</p>
---	---	--	--

## HTA Commissioning Board

**Chair,**  
**Professor Hywel Williams,**  
 Professor of Dermato-Epidemiology,  
 Centre of Evidence-Based Dermatology,  
 University of Nottingham

**Deputy Chair,**  
**Professor Jon Deeks,**  
 Department of Public Health and Epidemiology,  
 University of Birmingham

**Professor Tom Walley, CBE,**  
 Professor of Clinical Pharmacology,  
 Director, NIHR HTA programme,  
 University of Liverpool

### Members

<p>Professor Ann Ashburn,          Professor of Rehabilitation and Head of Research, Southampton General Hospital</p> <p>Professor Judith Bliss,          Director of ICR-Clinical Trials and Statistics Unit, The Institute of Cancer Research</p> <p>Professor Peter Brocklehurst,          Professor of Women's Health, Institute for Women's Health, University College London</p> <p>Professor David Fitzmaurice,          Professor of Primary Care Research, Department of Primary Care Clinical Sciences, University of Birmingham</p>	<p>Professor John W Gregory,          Professor in Paediatric Endocrinology, Department of Child Health, Wales School of Medicine, Cardiff University</p> <p>Professor Steve Halligan,          Professor of Gastrointestinal Radiology, University College Hospital, London</p> <p>Professor Angela Harden,          Professor of Community and Family Health, Institute for Health and Human Development, University of East London</p> <p>Dr Martin J Landray,          Reader in Epidemiology, Honorary Consultant Physician, Clinical Trial Service Unit, University of Oxford</p>	<p>Dr Joanne Lord,          Reader, Health Economics Research Group, Brunel University</p> <p>Professor Stephen Morris,          Professor of Health Economics, University College London, Research Department of Epidemiology and Public Health, University College London</p> <p>Professor Dion Morton,          Professor of Surgery, Academic Department of Surgery, University of Birmingham</p> <p>Professor Gail Mountain,          Professor of Health Services Research, Rehabilitation and Assistive Technologies Group, University of Sheffield</p>	<p>Professor Irwin Nazareth,          Professor of Primary Care and Head of Department, Department of Primary Care and Population Sciences, University College London</p> <p>Professor E Andrea Nelson,          Professor of Wound Healing and Director of Research, School of Healthcare, University of Leeds</p> <p>Professor John David Norrie,          Chair in Clinical Trials and Biostatistics, Robertson Centre for Biostatistics, University of Glasgow</p> <p>Dr Rafael Perera,          Lecturer in Medical Statistics, Department of Primary Health Care, University of Oxford</p>
--	---	--	--

## HTA Commissioning Board *(continued)*

Professor Barney Reeves,  
Professorial Research Fellow  
in Health Services Research,  
Department of Clinical Science,  
University of Bristol

Professor Peter Tyrer,  
Professor of Community  
Psychiatry, Centre for Mental  
Health, Imperial College London

Professor Martin Underwood,  
Professor of Primary Care  
Research, Warwick Medical  
School, University of Warwick

Professor Caroline Watkins,  
Professor of Stroke and Older  
People's Care, Chair of UK  
Forum for Stroke Training, Stroke  
Practice Research Unit, University  
of Central Lancashire

Dr Duncan Young,  
Senior Clinical Lecturer and  
Consultant, Nuffield Department  
of Anaesthetics, University of  
Oxford

### Observers

Dr Tom Foulks,  
Medical Research Council

Dr Kay Pattison,  
Senior NIHR Programme  
Manager, Department of Health

## HTA Clinical Evaluation and Trials Board

### Chair,

**Professor Sallie Lamb,**  
Director,  
Warwick Clinical Trials Unit,  
Warwick Medical School,  
University of Warwick and Professor of  
Rehabilitation,  
Nuffield Department of Orthopaedic,  
Rheumatology and Musculoskeletal Sciences,  
University of Oxford

### Deputy Chair,

**Professor Jenny Hewison,**  
Professor of the Psychology of Health Care,  
Leeds Institute of Health Sciences,  
University of Leeds

### Programme Director,

**Professor Tom Walley, CBE,**  
Director, NIHR HTA programme,  
Professor of Clinical Pharmacology,  
University of Liverpool

### Members

Professor Keith Abrams,  
Professor of Medical Statistics,  
Department of Health Sciences,  
University of Leicester

Professor Martin Bland,  
Professor of Health Statistics,  
Department of Health Sciences,  
University of York

Professor Jane Blazeby,  
Professor of Surgery and  
Consultant Upper GI Surgeon,  
Department of Social Medicine,  
University of Bristol

Professor Julia M Brown,  
Director, Clinical Trials Research  
Unit, University of Leeds

Professor Alistair Burns,  
Professor of Old Age Psychiatry,  
Psychiatry Research Group, School  
of Community-Based Medicine,  
The University of Manchester &  
National Clinical Director for  
Dementia, Department of Health

Dr Jennifer Burr,  
Director, Centre for Healthcare  
Randomised trials (CHART),  
University of Aberdeen

Professor Linda Davies,  
Professor of Health Economics,  
Health Sciences Research Group,  
University of Manchester

Professor Simon Gilbody,  
Prof of Psych Medicine and Health  
Services Research, Department of  
Health Sciences, University of York

Professor Steven Goodacre,  
Professor and Consultant in  
Emergency Medicine, School of  
Health and Related Research,  
University of Sheffield

Professor Dyfrig Hughes,  
Professor of Pharmacoeconomics,  
Centre for Economics and Policy  
in Health, Institute of Medical  
and Social Care Research, Bangor  
University

Professor Paul Jones,  
Professor of Respiratory Medicine,  
Department of Cardiac and  
Vascular Science, St George's  
Hospital Medical School,  
University of London

Professor Khalid Khan,  
Professor of Women's Health and  
Clinical Epidemiology, Barts and  
the London School of Medicine,  
Queen Mary, University of London

Professor Richard J McManus,  
Professor of Primary Care  
Cardiovascular Research, Primary  
Care Clinical Sciences Building,  
University of Birmingham

Professor Helen Rodgers,  
Professor of Stroke Care, Institute  
for Ageing and Health, Newcastle  
University

Professor Ken Stein,  
Professor of Public Health,  
Peninsula Technology Assessment  
Group, Peninsula College  
of Medicine and Dentistry,  
Universities of Exeter and  
Plymouth

Professor Jonathan Sterne,  
Professor of Medical Statistics  
and Epidemiology, Department  
of Social Medicine, University of  
Bristol

Mr Andy Vail,  
Senior Lecturer, Health Sciences  
Research Group, University of  
Manchester

Professor Clare Wilkinson,  
Professor of General Practice and  
Director of Research North Wales  
Clinical School, Department of  
Primary Care and Public Health,  
Cardiff University

Dr Ian B Wilkinson,  
Senior Lecturer and Honorary  
Consultant, Clinical Pharmacology  
Unit, Department of Medicine,  
University of Cambridge

### Observers

Ms Kate Law,  
Director of Clinical Trials,  
Cancer Research UK

Dr Morven Roberts,  
Clinical Trials Manager, Health  
Services and Public Health  
Services Board, Medical Research  
Council

## Diagnostic Technologies and Screening Panel

### Members

<p><b>Chair,</b> <b>Professor Lindsay Wilson Turnbull,</b> Scientific Director of the Centre for Magnetic Resonance Investigations and YCR Professor of Radiology, Hull Royal Infirmary</p> <p>Professor Judith E Adams, Consultant Radiologist, Manchester Royal Infirmary, Central Manchester &amp; Manchester Children's University Hospitals NHS Trust, and Professor of Diagnostic Radiology, University of Manchester</p> <p>Mr Angus S Arunkalaivanan, Honorary Senior Lecturer, University of Birmingham and Consultant Urogynaecologist and Obstetrician, City Hospital, Birmingham</p> <p>Dr Diana Baralle, Consultant and Senior Lecturer in Clinical Genetics, University of Southampton</p>	<p>Dr Stephanie Dancer, Consultant Microbiologist, Hairmyres Hospital, East Kilbride</p> <p>Dr Diane Eccles, Professor of Cancer Genetics, Wessex Clinical Genetics Service, Princess Anne Hospital</p> <p>Dr Trevor Friedman, Consultant Liaison Psychiatrist, Brandon Unit, Leicester General Hospital</p> <p>Dr Ron Gray, Consultant, National Perinatal Epidemiology Unit, Institute of Health Sciences, University of Oxford</p> <p>Professor Paul D Griffiths, Professor of Radiology, Academic Unit of Radiology, University of Sheffield</p> <p>Mr Martin Hooper, Public contributor</p>	<p>Professor Anthony Robert Kendrick, Associate Dean for Clinical Research and Professor of Primary Medical Care, University of Southampton</p> <p>Dr Nicola Lennard, Senior Medical Officer, MHRA</p> <p>Dr Anne Mackie, Director of Programmes, UK National Screening Committee, London</p> <p>Mr David Mathew, Public contributor</p> <p>Dr Michael Millar, Consultant Senior Lecturer in Microbiology, Department of Pathology &amp; Microbiology, Barts and The London NHS Trust, Royal London Hospital</p> <p>Mrs Una Rennard, Public contributor</p>	<p>Dr Stuart Smellie, Consultant in Clinical Pathology, Bishop Auckland General Hospital</p> <p>Ms Jane Smith, Consultant Ultrasound Practitioner, Leeds Teaching Hospital NHS Trust, Leeds</p> <p>Dr Allison Streetly, Programme Director, NHS Sickle Cell and Thalassaemia Screening Programme, King's College School of Medicine</p> <p>Dr Matthew Thompson, Senior Clinical Scientist and GP, Department of Primary Health Care, University of Oxford</p> <p>Dr Alan J Williams, Consultant Physician, General and Respiratory Medicine, The Royal Bournemouth Hospital</p>
--	--	---	---

### Observers

<p>Dr Tim Elliott, Team Leader, Cancer Screening, Department of Health</p> <p>Dr Joanna Jenkinson, Board Secretary, Neurosciences and Mental Health Board (NMHB), Medical Research Council</p>	<p>Professor Julietta Patnick, Director, NHS Cancer Screening Programme, Sheffield</p> <p>Dr Kay Pattison, Senior NIHR Programme Manager, Department of Health</p>	<p>Professor Tom Walley, CBE, Director, NIHR HTA programme, Professor of Clinical Pharmacology, University of Liverpool</p>	<p>Dr Ursula Wells, Principal Research Officer, Policy Research Programme, Department of Health</p>
--	--	---	---

## Disease Prevention Panel

### Members

<p><b>Chair,</b> <b>Professor Margaret Thorogood,</b> Professor of Epidemiology, University of Warwick Medical School, Coventry</p> <p>Dr Robert Cook, Clinical Programmes Director, Bazian Ltd, London</p> <p>Dr Colin Greaves, Senior Research Fellow, Peninsula Medical School (Primary Care)</p> <p>Mr Michael Head, Public contributor</p>	<p>Professor Cathy Jackson, Professor of Primary Care Medicine, Bute Medical School, University of St Andrews</p> <p>Dr Russell Jago, Senior Lecturer in Exercise, Nutrition and Health, Centre for Sport, Exercise and Health, University of Bristol</p> <p>Dr Julie Mytton, Consultant in Child Public Health, NHS Bristol</p>	<p>Professor Irwin Nazareth, Professor of Primary Care and Director, Department of Primary Care and Population Sciences, University College London</p> <p>Dr Richard Richards, Assistant Director of Public Health, Derbyshire County Primary Care Trust</p> <p>Professor Ian Roberts, Professor of Epidemiology and Public Health, London School of Hygiene &amp; Tropical Medicine</p>	<p>Dr Kenneth Robertson, Consultant Paediatrician, Royal Hospital for Sick Children, Glasgow</p> <p>Dr Catherine Swann, Associate Director, Centre for Public Health Excellence, NICE</p> <p>Mrs Jean Thurston, Public contributor</p> <p>Professor David Weller, Head, School of Clinical Science and Community Health, University of Edinburgh</p>
---	--	--	--

### Observers

<p>Ms Christine McGuire, Research &amp; Development, Department of Health</p>	<p>Dr Kay Pattison, Senior NIHR Programme Manager, Department of Health</p>	<p>Professor Tom Walley, CBE, Director, NIHR HTA programme, Professor of Clinical Pharmacology, University of Liverpool</p>
---	---	---

## External Devices and Physical Therapies Panel

### Members

<b>Chair,</b> <b>Dr John Pounsford,</b> Consultant Physician North Bristol NHS Trust	Dr Dawn Carnes, Senior Research Fellow, Barts and the London School of Medicine and Dentistry	Dr Shaheen Hamdy, Clinical Senior Lecturer and Consultant Physician, University of Manchester	Mr Jim Reece, Public contributor
<b>Deputy Chair,</b> <b>Professor E Andrea Nelson,</b> Reader in Wound Healing and Director of Research, University of Leeds	Dr Emma Clark, Clinician Scientist Fellow & Cons. Rheumatologist, University of Bristol	Professor Christine Norton, Professor of Clinical Nursing Innovation, Bucks New University and Imperial College Healthcare NHS Trust	Professor Maria Stokes, Professor of Neuromusculoskeletal Rehabilitation, University of Southampton
Professor Bipin Bhakta, Charterhouse Professor in Rehabilitation Medicine, University of Leeds	Mrs Anthea De Barton-Watson, Public contributor	Dr Lorraine Pinnigton, Associate Professor in Rehabilitation, University of Nottingham	Dr Pippa Tyrrell, Senior Lecturer/Consultant, Salford Royal Foundation Hospitals' Trust and University of Manchester
Mrs Penny Calder, Public contributor	Professor Nadine Foster, Professor of Musculoskeletal Health in Primary Care Arthritis Research, Keele University	Dr Kate Radford, Senior Lecturer (Research), University of Central Lancashire	Dr Nefyn Williams, Clinical Senior Lecturer, Cardiff University

### Observers

Dr Kay Pattison, Senior NIHR Programme Manager, Department of Health	Dr Morven Roberts, Clinical Trials Manager, Health Services and Public Health Services Board, Medical Research Council	Professor Tom Walley, CBE, Director, NIHR HTA programme, Professor of Clinical Pharmacology, University of Liverpool	Dr Ursula Wells, Principal Research Officer, Policy Research Programme, Department of Health
--	--	--	---

## Interventional Procedures Panel

### Members

<b>Chair,</b> <b>Professor Jonathan Michaels,</b> Professor of Vascular Surgery, University of Sheffield	Mr Seumas Eckford, Consultant in Obstetrics & Gynaecology, North Devon District Hospital	Dr Fiona Lecky, Senior Lecturer/Honorary Consultant in Emergency Medicine, University of Manchester/Salford Royal Hospitals NHS Foundation Trust	Professor Jon Moss, Consultant Interventional Radiologist, North Glasgow Hospitals University NHS Trust
<b>Deputy Chair,</b> <b>Mr Michael Thomas,</b> Consultant Colorectal Surgeon, Bristol Royal Infirmary	Professor Sam Eljamel, Consultant Neurosurgeon, Ninewells Hospital and Medical School, Dundee	Dr Nadim Malik, Consultant Cardiologist/Honorary Lecturer, University of Manchester	Dr Simon Padley, Consultant Radiologist, Chelsea & Westminster Hospital
Mrs Isabel Boyer, Public contributor	Dr Adele Fielding, Senior Lecturer and Honorary Consultant in Haematology, University College London Medical School	Mr Hisham Mehanna, Consultant & Honorary Associate Professor, University Hospitals Coventry & Warwickshire NHS Trust	Dr Ashish Paul, Medical Director, Bedfordshire PCT
Mr Sankaran Chandra Sekharan, Consultant Surgeon, Breast Surgery, Colchester Hospital University NHS Foundation Trust	Dr Matthew Hatton, Consultant in Clinical Oncology, Sheffield Teaching Hospital Foundation Trust	Dr Jane Montgomery, Consultant in Anaesthetics and Critical Care, South Devon Healthcare NHS Foundation Trust	Dr Sarah Purdy, Consultant Senior Lecturer, University of Bristol
Professor Nicholas Clarke, Consultant Orthopaedic Surgeon, Southampton University Hospitals NHS Trust	Dr John Holden, General Practitioner, Garswood Surgery, Wigan		Dr Matthew Wilson, Consultant Anaesthetist, Sheffield Teaching Hospitals NHS Foundation Trust
Ms Leonie Cooke, Public contributor			Professor Yit Chiun Yang, Consultant Ophthalmologist, Royal Wolverhampton Hospitals NHS Trust

### Observers

Dr Kay Pattison, Senior NIHR Programme Manager, Department of Health	Dr Morven Roberts, Clinical Trials Manager, Health Services and Public Health Services Board, Medical Research Council	Professor Tom Walley, CBE, Director, NIHR HTA programme, Professor of Clinical Pharmacology, University of Liverpool	Dr Ursula Wells, Principal Research Officer, Policy Research Programme, Department of Health
--	--	--	---

## Pharmaceuticals Panel

### Members

<b>Chair,</b> <b>Professor Imti Choonara,</b> Professor in Child Health, University of Nottingham	Dr James Gray, Consultant Microbiologist, Department of Microbiology, Birmingham Children's Hospital NHS Foundation Trust	Dr Maria Kouimtzi, Pharmacy and Informatics Director, Global Clinical Solutions, Wiley-Blackwell	Mrs Katrina Simister, Assistant Director New Medicines, National Prescribing Centre, Liverpool
<b>Deputy Chair,</b> <b>Dr Yoon K Loke,</b> Senior Lecturer in Clinical Pharmacology, University of East Anglia	Dr Jurjees Hasan, Consultant in Medical Oncology, The Christie, Manchester	Professor Femi Oyeboode, Consultant Psychiatrist and Head of Department, University of Birmingham	Professor Donald Singer, Professor of Clinical Pharmacology and Therapeutics, Clinical Sciences Research Institute, CSB, University of Warwick Medical School
Dr Martin Ashton-Key, Medical Advisor, National Commissioning Group, NHS London	Dr Carl Heneghan, Deputy Director Centre for Evidence-Based Medicine and Clinical Lecturer, Department of Primary Health Care, University of Oxford	Dr Andrew Prentice, Senior Lecturer and Consultant Obstetrician and Gynaecologist, The Rosie Hospital, University of Cambridge	Mr David Symes, Public contributor
Dr Peter Elton, Director of Public Health, Bury Primary Care Trust	Dr Dyfrig Hughes, Reader in Pharmacoeconomics and Deputy Director, Centre for Economics and Policy in Health, IMSCaR, Bangor University	Ms Amanda Roberts, Public contributor	Dr Arnold Zermansky, General Practitioner, Senior Research Fellow, Pharmacy Practice and Medicines Management Group, Leeds University
Dr Ben Goldacre, Research Fellow, Epidemiology London School of Hygiene and Tropical Medicine		Dr Gillian Shepherd, Director, Health and Clinical Excellence, Merck Serono Ltd	

### Observers

Dr Kay Pattison, Senior NIHR Programme Manager, Department of Health	Dr Heike Weber, Programme Manager, Medical Research Council	Dr Ursula Wells, Principal Research Officer, Policy Research Programme, Department of Health
Mr Simon Reeve, Head of Clinical and Cost- Effectiveness, Medicines, Pharmacy and Industry Group, Department of Health	Professor Tom Walley, CBE, Director, NIHR HTA programme, Professor of Clinical Pharmacology, University of Liverpool	

## Psychological and Community Therapies Panel

### Members

<b>Chair,</b> <b>Professor Scott Weich,</b> Professor of Psychiatry, University of Warwick, Coventry	Mrs Val Carlill, Public contributor	Dr Jeremy J Murphy, Consultant Physician and Cardiologist, County Durham and Darlington Foundation Trust	Dr Paul Ramchandani, Senior Research Fellow/Cons. Child Psychiatrist, University of Oxford
<b>Deputy Chair,</b> <b>Dr Howard Ring,</b> Consultant & University Lecturer in Psychiatry, University of Cambridge	Dr Steve Cunningham, Consultant Respiratory Paediatrician, Lothian Health Board	Dr Richard Neal, Clinical Senior Lecturer in General Practice, Cardiff University	Dr Karen Roberts, Nurse/Consultant, Dunston Hill Hospital, Tyne and Wear
Professor Jane Barlow, Professor of Public Health in the Early Years, Health Sciences Research Institute, Warwick Medical School	Dr Anne Hesketh, Senior Clinical Lecturer in Speech and Language Therapy, University of Manchester	Mr John Needham, Public contributor	Dr Karim Saad, Consultant in Old Age Psychiatry, Coventry and Warwickshire Partnership Trust
Dr Sabyasachi Bhaumik, Consultant Psychiatrist, Leicestershire Partnership NHS Trust	Dr Peter Langdon, Senior Clinical Lecturer, School of Medicine, Health Policy and Practice, University of East Anglia	Ms Mary Nettle, Mental Health User Consultant	Dr Lesley Stockton, Lecturer, School of Health Sciences, University of Liverpool
	Dr Yann Lefeuvre, GP Partner, Burrage Road Surgery, London	Professor John Potter, Professor of Ageing and Stroke Medicine, University of East Anglia	Dr Simon Wright, GP Partner, Walkden Medical Centre, Manchester
		Dr Greta Rait, Senior Clinical Lecturer and General Practitioner, University College London	

### Observers

Dr Kay Pattison, Senior NIHR Programme Manager, Department of Health	Dr Morven Roberts, Clinical Trials Manager, Health Services and Public Health Services Board, Medical Research Council	Professor Tom Walley, CBE, Director, NIHR HTA programme, Professor of Clinical Pharmacology, University of Liverpool	Dr Ursula Wells, Principal Research Officer, Policy Research Programme, Department of Health
--	--	--	---





### **Feedback**

The HTA programme and the authors would like to know your views about this report.

The Correspondence Page on the HTA website ([www.hta.ac.uk](http://www.hta.ac.uk)) is a convenient way to publish your comments. If you prefer, you can send your comments to the address below, telling us whether you would like us to transfer them to the website.

***We look forward to hearing from you.***