Health Technology Assessment 2001; Vol. 5: No. 35

A systematic review of controlled trials of the effectiveness and cost-effectiveness of brief psychological treatments for depression

R Churchill
V Hunot
R Corney
M Knapp
H McGuire
A Tylee
S Wessely

Health Technology Assessment
NHS R&D HTA Programme

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

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

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

You can order HTA monographs from our Despatch Agents:

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

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

Contact details are as follows:

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

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

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

Payment methods

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

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

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

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

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

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

Contents

A systematic review of controlled trials of the effectiveness and cost-effectiveness of brief psychological treatments for depression

R Churchill[1]*
V Hunot[1]
R Corney[2]
M Knapp[1]
H McGuire[1]
A Tylee[1]
S Wessely[1]

[1] Institute of Psychiatry, King's College London, UK
[2] University of Greenwich, UK

* Corresponding author

Competing interests: none declared

Published September 2002

This report should be referenced as follows:

Churchill R, Hunot V, Corney R, Knapp M, McGuire H, Tylee A, *et al*. A systematic review of controlled trials of the effectiveness and cost-effectiveness of brief psychological treatments for depression. *Health Technol Assess* 2001;**5**(35).

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

NHS R&D HTA Programme

The NHS R&D Health Technology Assessment (HTA) Programme was set up in 1993 to ensure that high-quality research information on the costs, effectiveness and broader impact of health technologies is produced in the most efficient way for those who use, manage and provide care in the NHS.

Initially, six HTA panels (pharmaceuticals, acute sector, primary and community care, diagnostics and imaging, population screening, methodology) helped to set the research priorities for the HTA Programme. However, during the past few years there have been a number of changes in and around NHS R&D, such as the establishment of the National Institute for Clinical Excellence (NICE) and the creation of three new research programmes: Service Delivery and Organisation (SDO); New and Emerging Applications of Technology (NEAT); and the Methodology Programme.

This has meant that the HTA panels can now focus more explicitly on health technologies ('health technologies' are broadly defined to include all interventions used to promote health, prevent and treat disease, and improve rehabilitation and long-term care) rather than settings of care. Therefore the panel structure has been redefined and replaced by three new panels: Pharmaceuticals; Therapeutic Procedures (including devices and operations); and Diagnostic Technologies and Screening.

The HTA Programme will continue to commission both primary and secondary research. The HTA Commissioning Board, supported by the National Coordinating Centre for Health Technology Assessment (NCCHTA), will consider and advise the Programme Director on the best research projects to pursue in order to address the research priorities identified by the three HTA panels.

The research reported in this monograph was funded as project number 95/22/04.

The views expressed in this publication are those of the authors and not necessarily those of the HTA Programme or the Department of Health. The editors wish to emphasise that funding and publication of this research by the NHS should not be taken as implicit support for any recommendations made by the authors.

Criteria for inclusion in the HTA monograph series
Reports are published in the HTA monograph series if (1) they have resulted from work commissioned 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 editors and publisher have tried to ensure the accuracy of this report but do not accept liability for damages or losses arising from material published in this report. They would like to thank the referees for their constructive comments on the draft document.

ISSN 1366-5278

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

G

List of abbreviations

BDI	Beck Depression Inventory
BT	behavioural therapy
CBT	cognitive behavioural therapy
CCDANCTR	Cochrane Collaboration Depression, Anxiety and Neurosis Controlled Trials Register
CCT	controlled clinical trial
CI	confidence interval
CT	cognitive therapy
df	degrees of freedom
DSM	Diagnostic and Statistical Manual
HRSD	Hamilton Rating Scale for Depression
IPT	interpersonal therapy
ITT	intention-to-treat
LDACL	Lubin Depression Adjective Checklist*
MDD	major depressive disorder
MMPI	Minnesota Multiphasic Personality Inventory
MMPI-D	unidimensional depression scale for MMPI*
NA	not applicable*
NIMH	National Institute of Mental Health
OR	odds ratio
QRS	Quality Rating Scale
PDT	psychodynamic therapy
PES	Pleasant Events Schedule*
RCT	randomised controlled trial
RDC	Research Diagnostic Criteria*
RGSD	Raskin Global Scale for Depression
RR	relative risk
SADS	Schedule of Affective Disorders Scale*
SCL	Symptom Checklist
SD	standard deviation
SE	standard error
SMD	standardised mean difference
ST	supportive therapy
WMD	weighted mean difference
ZSRDS	Zung Self-rating Depression Scale*

* Used only in tables and appendices

Executive summary

Background

Depression is common and can result in considerable impairment causing distress to sufferers and their families. It is also of substantial cost to the NHS and the wider economy. Most depression is treated in primary care, where both pharmacological and psychological treatments are used. The provision of psychological treatments is increasing and the effectiveness and cost-effectiveness of these interventions for depression need to be demonstrated.

Objectives

(1) To conduct a systematic review and, where possible, a meta-analysis of all controlled clinical trials (CCTs) in which brief psychological treatments were compared with one another or treatment as usual in the treatment of depression.
(2) To describe the internal validity, statistical power and external validity of the identified trials.
(3) To compare the overall efficacy of all variants of brief psychological treatments with treatment as usual.
(4) To compare the efficacy of cognitive behavioural therapy (CBT) with treatment as usual, interpersonal therapy (IPT), psychodynamic therapy (PDT) and supportive therapy (ST).
(5) To compare the efficacy of IPT, PDT and ST with treatment as usual and with one another.
(6) To compare the efficacy of all variants of individual and group therapies.
(7) To summarise all available cost data from controlled trials of brief psychological treatments for depression.

Methods

Data sources

A wide range of electronic bibliographic and specialist databases were searched using a comprehensive search strategy as appropriate. Eleven psychiatry/psychology and three economics journals were handsearched. In addition, bibliographies from the texts of relevant trials and reviews, grey literature (e.g. conference proceedings and government documents) and dissertations were searched. Leading researchers in the field, members of the International Network of Agencies for HTA, health authorities, UK counselling organisations and psychology department heads were also contacted.

Study selection

Published/unpublished randomised controlled trials (RCTs) or CCTs comparing different forms of brief psychological treatments (described within an explicit psychological orientation completed within a time-limited framework of ≤ 20 sessions), or brief psychological treatments with treatment as usual were included. Trial participants could be males or females aged 16–65 years with a primary diagnosis of depression. Marital/couples and family therapy were excluded.

Data extraction and synthesis

Qualitative and quantitative data relating to internal and external validity, study power and outcomes were extracted using a standardised spreadsheet. Odds ratios and relative risks were calculated for recovery and dropout data. Based on calculated weighted or standardised mean differences, fixed- and/or random-effects models were used to pool the mean differences and mean change data. Clinical and methodological heterogeneity were explored through heterogeneity and sensitivity analyses, and, where possible, other sources of bias were investigated using funnel plots. Finally, the cost-effectiveness data were summarised.

Results

Patients receiving any variant of psychotherapy were significantly more likely to improve to a degree where they were no longer considered clinically depressed, exhibited significantly fewer symptoms post-treatment and experienced greater symptom reduction from baseline than those receiving treatment as usual. No differences in treatment discontinuation were observed.

Patients receiving CBT were significantly more likely than those receiving PDT, IPT or ST to

improve to a degree where they were no longer regarded as being clinically depressed. No group differences in post-treatment symptoms, symptom reduction from baseline or dropouts during treatment were suggested.

Patients receiving individual therapies were significantly more likely to improve to a degree where they were no longer considered clinically depressed and exhibited fewer symptoms post-treatment. No differences in dropouts between groups were demonstrated.

No differences were demonstrated between cognitive and behavioural interventions in post-treatment recovery and symptoms, symptom reduction from baseline or dropouts.

Patients receiving variants of CBT were significantly more likely than those receiving treatment as usual to improve to a degree where they were no longer regarded as being clinically depressed and exhibited significantly fewer symptoms post-treatment and greater symptom reduction from baseline. No differences in dropouts between groups were demonstrated.

The evidence comparing variants of CBT with IPT was limited, but suggested that there were no differences in post-treatment recovery and dropouts during treatment. Patients receiving variants of CBT were significantly more likely than those receiving PDT to improve to a degree where they were no longer regarded as being clinically depressed, although no group differences in post-treatment symptoms, symptom reduction from baseline or dropouts were suggested. Patients receiving variants of CBT were significantly more likely than those receiving ST to improve to a degree where they were no longer considered clinically depressed and exhibited fewer symptoms post-treatment. No group differences in symptom reduction from baseline or dropouts during treatment were demonstrated.

Patients receiving ST were significantly more likely than those receiving treatment as usual to improve to a degree where they were no longer considered clinically depressed and exhibited fewer symptoms post-treatment. No group differences in symptom reduction from baseline or dropouts were suggested.

Trials comparing IPT with ST, IPT with treatment as usual and PDT with ST all yielded insufficient data upon which to base any firm conclusions.

It was not possible to draw any firm conclusions from the limited follow-up and economic data available, although economic evidence provided tentative support for the hypothesis that psychotherapy was more efficient than usual care and suggested a modest cost-effectiveness advantage in favour of CBT.

Low overall quality scores were recorded for many of the trials. Methodological problems were noted relating to the randomisation and allocation procedures, exclusion of randomised patients, sample size, use of concurrent treatments, investigator bias, monitoring of therapist adherence and use of broader outcome measures (e.g. quality of life). Interpretation of the findings was further limited by the identification of probable bias in the funnel plots and heterogeneity and sensitivity analyses. Doubt exists as to the generalisability of the trials identified to UK primary care settings in terms of socio-demographic characteristics, severity of disorder, motivation of participants and therapy type.

Conclusions

Implications of the review for healthcare

Based on the best available evidence, it would appear that some forms of brief psychological treatments, particularly those derived from cognitive/behavioural models, are beneficial in the treatment of people with depression being managed outside hospital settings. Little can be said about the efficacy of different types of individual versus group therapy because all the trials comparing these formats used CT or BT. In these trials, greater efficacy for individual formats was suggested.

Baseline severity, the methods used to identify patients and possibly the number of sessions offered are factors likely to affect outcome. Little can be said about the potential impact of socio-demographic characteristics of patients, the specific effects of client motivation and therapeutic alliance, any potential adverse events associated with psychological treatments, the short- and long-term outcomes of psychological treatments, the differential effects of alternative models, particularly PDT and client-centred therapies, or the immediate and long-term economic consequences attached to the provision of psychological treatments in primary care.

Implications for research

Further trials of all types of psychological treatments in primary care settings involving appropriately recruited representative patient

samples, whose disorders have been recognised and who meet the diagnostic criteria for depressive disorder, are required. RCTs examining both immediate and long-term outcomes and cost implications and trials, both brief and long term, of PDT or client-centred therapies (using manualised/standardised techniques), and of different psychological treatments in individual versus group formats are required. Future trials need to be adequately powered, involve longer follow-up, properly monitor adherence to therapeutic technique, incorporate outcomes measuring the broader impact of treatment, provide adequately powered high-quality cost data and record and allow for the use of any non-randomised concomitant treatments.

Chapter 1

Background and objectives

The syndrome of depression

The term depression describes a broad spectrum of moods, behaviour and emotions ranging from the syndrome of major depressive disorder (MDD) to lifelong mild fluctuating depression (dysthymia) upon which a major depressive episode may be superimposed, and other less severe subclinical states.[1] Central features of depression are dysphoria, loss of interest in normally enjoyable activities, impaired concentration and memory difficulties and feelings of worthlessness, often associated with suicidal ideation. It may be accompanied by physical symptoms, such as sleep disturbance, weight change, fatigue and the somatic symptoms of anxiety. Depression can produce significant distress and impairment in many domains of functioning that may equal or exceed that of chronic medical conditions and, furthermore, may result in a major burden of suffering among patients and their families.[2]

Epidemiology and costs of depression

Depression is common, with a point prevalence of 15–30% in the UK adult population, depending upon the threshold used for case definition.[3] It is estimated that about 60–70% of adults will at some time in their lives experience depression or worry of sufficient severity to influence their daily activities,[4] and that approximately 18–26% of females and 8–12% of males will suffer from a major depressive episode.[5] More than 50% of patients who attend their general practice may present with symptoms of depression[6] and, furthermore, up to 50% of cases of depression remain undetected.[7]

The vast majority of depressed patients do not require hospitalisation and 90% of depression in the UK is managed in primary care.[4] However, depression is of considerable financial cost to the UK NHS.[8,9] The costs of treating depression have been shown to outweigh those of treating psychoses.[10] Kind and Sorensen[8] estimated the annual NHS costs to be £417 million (at 1991 prices) and the indirect costs of lost work days and premature mortality to be almost £3000 million. The picture is similar in the USA where the costs associated with lost work productivity in untreated depression have been estimated to be substantial.[11] Indeed, one study has suggested that depression costs the USA economy $44 million million annually, three-quarters of which is accounted for by the indirect costs.[12]

Psychological treatments for depression

There has been a noticeable increase in the number of patients seeking non-pharmacological approaches to alleviate symptoms of depression. The growth in demand for psychological treatments appears to have been motivated by concerns about medication side-effects, beliefs about potential drug dependency and reluctance to use a biochemical treatment in order to resolve psychological problems. An observed shift in public attitude towards psychotherapy and counselling as an acceptable and preferred intervention has been quantified in a survey of lay peoples' attitudes towards depression conducted by the Defeat Depression Campaign. This survey revealed that 91% of respondents thought that people with depression should be offered counselling, and just 16% thought that they should be offered antidepressants.[13] In response to these changes in treatment preference combined with low adherence rates to antidepressants,[1,14] failure of some patients to tolerate or respond to medication and contraindications to pharmacotherapy,[15] psychological treatments have become increasingly available as an alternative intervention for patients with depression within the NHS in recent years.

In contrast to 30 years ago when traditional insight-oriented psychotherapy tended to be the sole psychological treatment model offered to depressed patients, there are now a plethora of newly developed therapeutic approaches for the treatment of depression. Behavioural therapy (BT) became more widely available during the 1970s, from which social skills training became a manualised intervention (i.e. an intervention guided by a formal manual) for depression.[16] With the advent and concomitant expansion of cognitive therapy (CT),[17] 'hybrid' therapeutic

approaches incorporating features of both CT and BT (including techniques such as relaxation) have proliferated[18] and include problem-solving therapy, self-control therapy, assertion training and psychoeducational programmes. Other models that integrate elements of psychodynamic and cognitive approaches have been developed and standardised in recent years, of which interpersonal therapy (IPT) is the best known and most utilised in the field of depression.[19]

Many manualised psychotherapy models have been devised and operationalised for group therapy formats, in order to facilitate and potentiate the development of interpersonal relationships within a group process. In economic terms, group therapy costs would appear to be lower. The average cost for a single session of individual counselling is calculated to be £56 per hour of patient contact, including preparation and administration time and additional costs of supervision and support.[20] In contrast, in a group therapy session of 90 minutes, an average number of eight patients may be facilitated by one, or sometimes two, therapists.

Psychological treatments may be delivered by a wide range of healthcare practitioners in the UK, including psychologists, counsellors, practice nurses, health visitors, community psychiatric nurses and social workers.[21] Psychological services are currently offered in over 50% of primary care practices in the UK, and are usually undertaken by trained counsellors. According to training undertaken, personal preference and presenting problems of clients, counsellors may choose to utilise a single psychotherapeutic model or an integrated combination of models. Preliminary findings from a survey of primary care counsellors conducted at the beginning of 1998 revealed that 42% of respondents described their approach as integrative when working with clients who present with a primary diagnosis of depression. Other models utilised for depression by counsellors in the survey were psychodynamic (20%), person-centred (21%) and cognitive behavioural (11%).[22]

One of the principal features of psychological treatments currently practiced in healthcare settings is the defined number of sessions offered. Time-limited psychotherapy has been increasingly developed and utilised partly as a response to economic and funding considerations[23,24] and partly because of patient preference for brief intervention.[25,26] There appears to be consensus in the literature that up to 20 sessions in all models of psychotherapy constitutes a time-limited therapeutic framework.[27,28] Recently developed psychotherapy models, such as CT and IPT, were specifically formulated as time-limited interventions in the treatment of depression, and the psychodynamic approach has also developed brief therapy models in response to the requirement for time-limited therapy. The efficacy of brief intervention has been demonstrated by Howard and colleagues,[29] who used a probit analysis model to conduct a meta-analysis to provide estimates of expected benefits of specific 'doses' of psychotherapy. Fifteen sets of data that had been gathered before 1969 were included in the analysis, involving psychodynamic and interpersonal models of psychotherapy that were available at that time. The results showed a dose–effect curve for the number of psychotherapy sessions indicating that 62% of patients were likely to have improved by the 13th session, after which there appeared to be diminishing returns. Further evidence of a lack of differences in outcomes between short and longer-term therapies was indicated in a meta-analysis conducted by Smith and Glass.[30] Furthermore, lower attrition rates in time-limited therapy have been demonstrated when compared with open-ended and long-term treatment contracts.[31]

Studies of psychotherapy for depression

Randomised controlled trials (RCTs) are properly regarded as the preferred approach to studying the efficacy of therapeutic interventions. Eligible subjects are randomly allocated to groups receiving one or more experimental interventions or a control intervention (which may be placebo or 'usual treatment'). Observations (often measures of clinical outcomes) from the different groups are compared, usually by an operator 'blinded' to the group to which the patient (or subject) was allocated. (Preferably, subjects are also blinded to the group to which they have been allocated.) These study design features of randomisation, comparison with a control and blinding of subjects and operator minimise potential biases that may otherwise affect study results.

Certain challenges are peculiar or particularly relevant to the design of RCTs of psychological treatments. Firstly, it is not generally feasible to design a true placebo control, and this limits comparisons to 'treatment as usual' or 'no treatment' (or waiting list), or comparisons between different forms of psychological treatment. Secondly, it is difficult or impossible to blind

the subject to whether they are receiving the intervention under study or a control intervention, and it may be difficult to blind the operator taking observations. Thirdly, other factors, such as therapist experience, treatment delivery, subject selection and assessment procedures, may be difficult to control and may, therefore, confound results.

Individual trials of psychological treatments for depression fail to demonstrate consistent superior outcomes in favour of any single therapeutic model. This may be largely attributable to non-specific treatment effects fundamental to all these therapies. For example, Hollon[32] defines a group of non-specific strategies used by therapists, such as receptive listening, professional manner, warmth, empathy, involvement/interest, genuineness and rapport, which might be expected to have such effects. McLean[33] has identified a number of features that BT, CT and IPT describe as specific techniques, but which are common to all three models. These include collaborative treatment process, focus on the present and future, proactive therapist intervention, time-limited application and emphasis on structure in the treatment programme. This lack of treatment specificity may limit the potential for demonstrating differences in efficacy between models.

Most outcome studies of psychological treatments seek to assess the efficacy of cognitive and behavioural models. There are fewer trials of psychodynamic therapy (PDT) or client-centred therapy, and these are frequently utilised as attention-control placebos in the trials of CTs and BTs, suggesting possible researcher bias or 'allegiance' in favour of these other models. It has been suggested that this could account for some inconsistencies between studies comparing different psychotherapeutic models.[34,35] It should also be noted that there are a number of trials in which psychotherapy interventions are undertaken in totality by the experimenter, with the potential of bias towards the active therapeutic intervention, blurring of model differences due to replication of therapist characteristics, contamination and possible subjective interpretation of results.

Although the need for economic appraisal is now firmly established, comparatively few evaluations have included cost-effectiveness or other economic dimensions to date,[36] and, in the majority of these studies, the measurement of costs and economic effects has not been comprehensive.[37] Previous research has sometimes reported direct healthcare costs and, occasionally, the effects of treatment on labour market participation by the patient, but other costs and wider economic effects have tended to be overlooked.

Rationale for a systematic review and meta-analysis of RCTs in the psychological treatment of depression

A systematic review is prepared using a systematic and rigorous approach to minimising biases and random errors. Systematic reviews involve a clear *a priori* description of their objectives, the methods to be used to identify primary studies, the criteria for inclusion or exclusion of studies, the methods to be used for assessing the methodological quality of included studies and the methods to be used for pooling data. Meta-analyses involve the statistical pooling of results from a group of independent primary studies into a single estimate of the intervention effect. The technique offers increased statistical power for assessing the impact of an intervention and increases the precision of the overall estimate of effect.

Previous reviews in the field of psychological treatments and depression have tended to be narrative and non-systematic and have thus been prone to several types of bias, which limit their use. More recent systematic reviews have set broad diagnostic inclusion criteria, such as psychological distress or common mental disorders, or have been limited by inclusion criteria in terms of psychological treatments, outcome measures or settings in which the psychotherapy is conducted. Many reviews are now outdated (for example, a comprehensive review and meta-analysis conducted by Robinson and colleagues in 1990[34] does not include trials carried out after 1987), have failed to use a thorough and systematic search strategy or have not conducted a meta-analysis of quantitative data.

This review will appraise and summarise all controlled trials of psychological treatments for depression in outpatient settings. A meta-analysis will be conducted where possible in order to establish whether there is statistical evidence that brief psychological treatments provide an effective intervention for depression measured in terms of significant clinical improvement, mean differences and mean change at post-treatment and, where possible, at follow-up. Comparisons will be conducted between different psychological models to identify differences in their treatment effects.

Relevant economic evidence reported in the literature will also be summarised to assess whether brief psychological interventions offer a cost-effective option for the treatment of depression.

Objectives

(1) To conduct a systematic review and, where possible, a meta-analysis of all controlled trials in which brief psychological treatments are compared with one another, or with treatment as usual/waiting-list conditions, in the treatment of depression.
(2) To describe the internal validity, statistical power and external validity of the identified trials.
(3) To compare the overall efficacy of all variants of brief psychological treatments with treatment as usual/a waiting-list control, at post-treatment and, where possible, at follow-up.
(4) To compare the efficacy of cognitive behavioural therapy (CBT) with treatment as usual/a waiting-list control, and with IPT, PDT and supportive therapy (ST) at post-treatment and, where possible, at follow-up.
(5) To compare the efficacy of IPT, PDT and ST with treatment as usual/a waiting-list control, and with one another at post-treatment and, where possible, at follow-up.
(6) To compare the efficacy of all variants of individual and group therapy.
(7) To summarise all available data from controlled clinical trials (CCTs) on the cost-effectiveness of brief psychological treatments for depression.

Chapter 2
Methods

Inclusion criteria for the review

Studies
Studies eligible for inclusion were RCTs or CCTs, both published and unpublished. All trials ever undertaken in any country were eligible for inclusion.

Participants
Trial participants could be males or females aged 16–65 whose primary diagnosis was depression according to the Research Diagnostic Criteria,[38] the Diagnostic and Statistical Manual (DSM)-IIIR/IV,[39] the International Classification of Disease[40] criteria or other validated diagnostic instruments, or who were assessed for levels of depressive symptomatology through self-rated or clinician-rated validated instruments.

Interventions
Interventions eligible for inclusion included all psychotherapies that were described within an explicit psychological orientation, and were completed within a time-limited framework of ≤ 20 sessions. The psychological treatments had to be compared with treatment as usual, or with each another, and both individual and group psychotherapies were eligible for inclusion.

Marital/couples therapy was excluded from the review because this format of intervention involved the partners of depressed subjects in the therapeutic intervention, with the relationship between couples utilised as the primary focus rather than the depression. A systematic review to examine the effectiveness of marital therapy for depression will be conducted separately.

Outcome measures
The main outcome measure was depression symptom level. Symptom levels had to be measured using self-rating scales, such as the Beck Depression Inventory (BDI)[41] and the Minnesota Multiphasic Personality Inventory (MMPI),[42] and/or clinician-rating scales, such as the Hamilton Rating Scale for Depression (HRSD).[43] Due to the paucity of economic analyses in this area, trials did not have to incorporate cost-effectiveness data in order to be included.

Search strategy for the identification of studies

Electronic bibliographic databases
The electronic search strategy was modified and refined several times and the following electronic bibliographic databases were searched: MEDLINE, PsycInfo, EMBASE, Science Scisearch and Social Scisearch. The Cochrane Collaboration Depression, Anxiety and Neurosis Controlled Trials Register (CCDANCTR – containing 11,050 references to RCTs/CCTs in depression, anxiety and neurosis as at March 1999) and those of other Cochrane groups were also comprehensively searched. Three updating searches of the CCDANCTR were carried out at quarterly intervals over the 12-month study period. The search strategies used are listed in appendix 1.

Handsearching
Eleven psychiatry and psychology journals (*Journal of Psychosomatic Research*, *Journal of Behavioral Medicine*, *Psychotherapy and Psychosomatics*, *British Journal of Social and Clinical Psychology*, *Psychotherapy*, *Clinical Psychology and Psychotherapy*, *Psychotherapy: Theory, Research and Practice*, *Journal of Counseling Psychology*, *Clinical Psychology Review*, *Comprehensive Psychiatry* and *International Journal of Group Psychotherapy*) identified as being likely to contain relevant RCTs were handsearched to identify references to randomisation procedure within the text. The selected articles were then examined to establish relevance to this review.

Three health economics journals (*Journal of Mental Health Policy and Economics*, *Journal of Health Economics* and *Health Economics*) identified as being likely to contain relevant RCTs with economic and/or cost-effectiveness dimensions were also handsearched.

Reference lists
References and bibliographies from the text of reports of relevant trials and reviews were examined for further RCTs not previously identified, and for papers relating to economic analyses.

Key researchers
Leading researchers in the field in the UK and USA were contacted by letter with a list of

the inclusion criteria for the review, and were asked for information regarding any additional published and unpublished trials.

Grey literature

Grey literature, such as conference abstracts/proceedings, government documents and other literature outside of the main journal literature, were identified and handsearched where possible. The following specialist databases were also searched: System for Information on Grey Literature, National Technical Information Service, DHSS-Data and the British Reports and translations and theses received by the British Library Document Supply Centre.

Theses and dissertations

Theses and dissertations were identified through reference and bibliography lists. The Dissertation Abstracts International database was also searched through PsycINFO, using the comprehensive search strategy developed for the other electronic database searches.

HTA reports

Other *HTA* reports of systematic reviews available via the NHS Centre for Reviews and Dissemination Database of Abstracts of Reviews of Effectiveness were searched. Members of the International Network of Agencies for HTA were also contacted in the hope of identifying research commissioned in this area but not published in English.

NHS National Research Register

The NHS National Research Register was examined for information on current and recent research projects. This register was searched for all studies containing the words 'depression'/'depress*' and 'psychotherapy'/'psychotherap*'. Further details were obtained for all references located in this way and these were cross-referenced with studies identified from other sources to ensure comprehensive cover.

Health authorities

All health authorities and health agencies in the UK were contacted by letter and by email to obtain reports of any relevant studies conducted in their area, or reports on informal evaluations of these interventions as part of audit procedure.

University psychology departments

All psychology department heads in the UK were contacted by letter to locate eligible unpublished or ongoing studies.

Counselling and psychotherapy organisations

Organisations involved in providing counselling and psychological treatments were approached for information through the individual and organisational members of the British Association for Counselling Research Network. Details of the search for eligible studies were circulated around the British Psychological Society Divisions of Clinical Psychology and Counselling Psychology and a letter requesting information on unpublished studies was placed in *The Psychologist*, a journal sent to all members of the British Psychological Society. In addition, a brief report about the review, incorporating a request for unpublished studies, was published in *Counselling Journal*.

Methods of review

Qualitative and quantitative data were extracted relating to the internal validity, study power and external validity. A standardised data extraction sheet was completed for every included trial, which recorded information on the study population, interventions, randomisation and blinding procedures, sample size, outcome data, follow-up and statistical analysis. A spreadsheet was constructed for entering extracted data from each trial to provide a detailed descriptive analysis (see appendix 2).

Obtaining unpublished data for the included trials

Attempts were made to obtain data that had not been reported in published studies, by contacting the first author of each study through dissemination of a standard letter that explained the purposes of the review and specified the additional data required (see appendix 3). A follow-up letter was sent to non-responding authors 1 month after the initial letter had been sent out. Where no further usable data were provided, studies were not included in the meta-analysis, and were listed as excluded due to missing data.

Methods of analysis

Treatment comparisons

A diagrammatic representation of all psychotherapy models utilised in the included trials was developed in order to identify the theoretical derivation of each model (see appendix 4). Each model was placed in a defined category based on

the authors' description. A table of comparisons was constructed to compare categorised psychotherapy models with treatment as usual control conditions and with one another, and to compare individual versus group therapy. The treatment as usual control condition included usual care/management, waiting list and no treatment.

Treatment outcomes

The main outcomes used by trials were symptom levels measured by rating scales presented as continuous (means and standard deviations (SDs)), categorical (recovery/improvement/no change) or dichotomous outcomes (recovery/non-recovery). Quality-of-life outcomes as measured by social and cognitive behavioural functioning (including self-esteem and assertiveness scales) provided additional outcomes where available. Economic outcomes transformed from these measures were also utilised. Cost-effectiveness data were extracted where possible, based on cost measures of varying breadth and (usually) on a summary outcome indicator. Dropouts from psychotherapy were used as a secondary outcome.

Choice of methods for pooling data

Due to the fact that the populations of patients from the individual trials in these analyses were different, ratios rather than absolute differences were calculated. In the first instance, dichotomous data were pooled using odds ratios (ORs), that is, the ratio of the odds of an event in the intervention group to the odds of an event in the control group. Odds are the ratio of the number of people in a group with an event to the number without an event. An OR of 1.0 indicates no difference between comparison groups. For undesirable outcomes, an OR of < 1.0 indicates that the intervention was effective in reducing the risk of that outcome. In this report, a reduction of or relief from depressive symptoms and retention of subjects in treatment are the desirable outcomes. An OR of > 1.0 indicates the efficacy of an 'experimental' intervention versus the comparison. When the event rate is small, ORs are very similar to relative risks (RRs). RRs were also calculated and reported in the text.

Two types of continuous data were extracted and pooled. The mean change from baseline was defined as the difference in mean change between the intervention and comparison groups divided by the SDs of both groups. Mean difference was defined as the difference in means between the intervention and comparison groups divided by the SDs of both groups. Two methods for pooling continuous data were used. Where all trials measured outcomes on the same scales and where the mean, SD and sample size in each group were known, weighted mean differences (WMDs) were calculated. The weight given to each study was determined by the precision of its estimate of effect. In the statistical software used to analyse the data in this review, Review Manager, this is equal to the inverse of the variance. Where some of the trials measured outcomes on different scales and it was not considered appropriate to directly combine data from these measures, standardised mean differences (SMDs), that is, the difference between two means divided by an estimate of the within-group SD, were calculated.

Both dichotomous and continuous outcomes are presented with 95% confidence intervals (CIs). These provide the range within which the 'true' value (e.g. size of effect of an intervention) is expected to lie with 95% certainty.

Where there was no evidence of statistical heterogeneity, a fixed-effects model was used in the first instance to combine data. This statistical model stipulates that the units under analysis (e.g. people in a study in a meta-analysis) are the ones of interest, and thus constitute the entire population of units. Only within-study variation is taken to influence the uncertainty of results (as reflected in the CI) of a meta-analysis using a fixed-effects model. Variation between the estimates of effect from each study (heterogeneity) does not affect the CI in a fixed-effects model. Where there was evidence of statistical heterogeneity, results were recalculated using a random-effects model. In this statistical model, both within-study sampling error (variance) and between-study variation are included in the assessment of the uncertainty (CI) of the results of a meta-analysis. If there is significant heterogeneity among the results of the included studies, random-effects models will give wider CIs than fixed-effects models.

Heterogeneity

Heterogeneity refers to variability or differences between studies in the estimates of effects. A distinction was made between 'statistical heterogeneity' (differences in the reported effects), 'methodological heterogeneity' (differences in study design) and 'clinical heterogeneity' (differences between studies in key characteristics of the participants, interventions or outcome measures). A formal test for statistical heterogeneity, the natural approximate χ^2 test, was conducted to assess whether the observed variability in study results (effect sizes) was greater than that expected to occur by chance. Clinical heterogeneity was explored according to two characteristics specified

a priori, namely the severity of depression at baseline and the number of psychotherapy sessions offered. Severity of depression was divided into three categories: severe, mild/moderate and unspecified. The number of psychotherapy sessions offered was divided, where possible, into three categories: one to six, seven to 12 and 13 to 20 sessions. Exploration of methodological heterogeneity was undertaken through sensitivity analyses as described below.

Sensitivity analysis

Sensitivity analysis provides an approach for testing how robust the results of a review are relative to key decisions and assumptions that have been made in the process of conducting the review. Sensitivity analyses were used to examine the impact of two aspects of methodological heterogeneity on the results of the review, one specified *a priori* and one *post hoc*. The impact of including studies of variable quality in terms of their design and methodological characteristics was specified *a priori*. Inclusion of lower-quality studies was thought likely to have an impact on the results of the review. Thus, sensitivity analyses were carried out where appropriate to investigate the influence of lower-quality studies on the results of the review. Where possible, studies were divided into four categories, namely scores of 0 to 9 (lowest quality), 10 to 19, 20 to 29 and 30 or more (highest quality), and the quality scores were ascertained as described below.

This review was intended to inform the clinical management of depressed patients in primary care settings. Due to the varying methods by which patients were recruited into the trials, it was decided *post hoc* to explore the effects of the method by which patients were recruited on the results of the review. This seemed the most appropriate proxy marker for the type of patients likely to present with clinically significant and recognised symptoms in primary care settings. Trials were divided into two categories, namely those trials that recruited from primary care and outpatient clinics or that took referrals from treating clinicians, and those that recruited patients through advertisements, used volunteers or used a combination of these two methods. In addition, the possibility of alternative sources of bias was investigated using funnel plots.

Investigation for other sources of bias

Where sufficient numbers of trials allowed a meaningful presentation, funnel plots were constructed to establish whether other potential biases could be operating. Funnel plots provide a graphical display of sample size plotted against effect size that can be used to investigate publication bias. When many studies have been located that estimate the same effect, the distribution of points should resemble a funnel shape with a widening in the spread of effect sizes as sample size decreases. A gap on one side of the wide part of the funnel indicates that some studies have not been published or located.

Methodological quality of included studies

Assessments of the quality of included trials

A simple but thorough approach was used to assess the quality and generalisability of the included studies. The overall quality of the trials was scored according to important elements of design and conduct. Study quality was examined using 23 criteria, and each item was scored between 0 and 2 with a maximum total score of 46. See appendix 5 for the Quality Rating Scale (QRS) used to assess and give a score for the overall quality of the studies.[44] The total scores of each trial were then categorised and used as an estimate of trial quality in conducting the sensitivity analyses.

The 23 items were grouped into four categories relating to different aspects of trial design. Total scores for each study were calculated in the four categories. Assessments of the internal validity of the trials could then be made to establish whether each trial was a fair comparison of the specific treatments studied on the specific patients recruited. The scores from grouped items also enabled an assessment of the external validity or generalisability of each trial to establish the relevance of the trials in determining treatment policy for patients.

Chapter 3
Results

Description of studies

Studies included

A total of 63 relevant studies were identified through the following sources: 35 from reference/bibliography lists, 24 from electronic bibliographic databases, two from known investigators in the field, one from handsearching and one from an unpublished thesis. The included trials were all published in English between 1973 and 1998. See appendix 6 for a list of the references and study numbers relating to the included trials.

Studies excluded

A further 13 studies were identified but excluded: two gave further medical or exercise interventions to the waiting-list control group, two utilised a crossover treatment design, four included patients that did not meet the inclusion criteria for the review, two evaluated different components of the same treatment, two described an educational rather than a psychotherapeutic intervention and one manipulated their randomisation procedure in accordance with patient characteristics (see appendix 7 for a list of the references of the trials that were excluded, together with reasons for exclusion).

Tables 1–9 summarise the basic characteristics of the included trials in terms of the populations studied, the treatments examined, the actual application of psychotherapy, the outcome measures used, the randomisation, allocation concealment and blinding procedures applied, patient follow-up, approaches to statistical analysis, treatment outcomes and whether antidepressant medication was used. The data extraction spreadsheet used to record this information for each trial is presented in appendix 2. A more detailed table of patient characteristics and interventions for each trial is presented in appendix 8.

Primary and secondary outcomes were measured using a total of 79 validated and referenced instruments across all the included trials (see appendix 9). The most frequently utilised instruments are presented in *Table 5*. All trials used more than one outcome measure and, therefore, percentages do not total 100%.

The main outcomes used by trials were symptom levels measured by rating scales presented as continuous or dichotomous outcomes (*Table 9*). Some trials reported both continuous and dichotomous outcomes at post-treatment assessment and final follow-up and, therefore, percentages in this table do not total 100%. A total of 50 trials reported sufficient data for inclusion in the meta-analysis as detailed in *Table 9*.

Unpublished data from included trials

All trials had missing data and 51 authors (six authors had conducted two or more included trials) were contacted by letter to request unreported information. Authors who failed to reply were sent a follow-up letter 1 month later. Of those contacted, 26 authors or their colleagues responded (51% response rate). Of these, ten authors provided new information (19.6% of all authors contacted), four had destroyed or mislaid relevant data sets, two were deceased and ten did not provide the requested information.

Sample size

The number of subjects, adequacy of follow-up and reported precision of the estimate of effect for each trial were recorded to examine the role of sample size. The majority of identified studies were small. The total sample size ranged from 18–276 subjects, and the median total sample size was 44 subjects (interquartile range 32–89). Five studies failed to report group sizes at randomisation (studies 1, 10, 24, 56 and 60*). The remaining 58 studies had a median group size of 13 subjects, and group sizes ranged from a minimum of six to a maximum of 109. Using figures taken from existing literature in the field, calculations were made using the epidemiological statistics software Epi-Info–Version 6 to ascertain the sample size that might be appropriate for a trial of this nature. These are presented for comparison purposes as follows.

* See appendix 6 for the references for each study.

(a) Psychotherapy (any variant) versus treatment as usual

In general, approximately 50% of subjects participating in any variant of brief psychological treatment will have 'recovered' at the end of treatment.[45] McLean and colleagues[33] reported that 20% of subjects recover at the end of treatment due to course effects, that is, spontaneous improvement without intervention. With the significance level set at 5%, 57 subjects would be required in each group to observe a significant treatment difference with 90% power. From a total of 30 trials that compared a model of psychotherapy with treatment as usual (studies 2, 3, 6–9, 14, 16, 19, 21, 22, 24, 25, 30, 33–35, 38, 40, 43–46, 48, 50, 51, 53, 54, 58 and 60*), two trials (studies 25 and 45*) had group sizes of sufficient power to detect a statistically significant difference.

(b) Psychotherapy (model 1) versus psychotherapy (model 2)

In general, 50% of subjects participating in any form of psychological treatment will be said to have recovered by the end of treatment. In addition, 60% of gain in symptom improvement may be attributable to a combination of treatment effects and course effects,[45] of which course effects may account for approximately one-third. Therefore, 40% of the gain in symptom improvement may be attributed to non-specific treatment effects. With the significance level set at 5%, 538 subjects would be required in each psychotherapy group to observe a significant treatment difference between two psychotherapy models with 90% power. With 80% power, 407 subjects would be required in each sample group to detect a treatment difference. From a total of 32 trials that compared two or more models of psychotherapy (studies 1, 4, 5, 8, 10–13, 15–18, 22, 23, 26–32, 36, 39, 42, 45, 49, 50, 52, 53, 56, 60 and 61*), none of them had treatment group sizes of sufficient power to detect a difference.

Quality rating of trials

From a possible maximum total score of 46 on the QRS, the mean overall quality score attained by the included trials was 19 (SD = 7.03, range 7–35). See appendix 10 for the score of each of the 23 quality items for each included study. As shown in *Tables 10* and *11*, the 23 items of the QRS were grouped according to their relevance to different aspects of trial design, as follows:

(1) subjects/patients
(2) psychological intervention
(3) internal validity (i.e. trial design/ scientific rigour)
(4) external validity (i.e. generalisability).

Model groupings

From the included trials, 32 distinct psychological models or psychotherapeutic techniques for the treatment of depression were identified. Through descriptive information about the models and references provided by authors, it was possible to locate the theoretical origins of each model under examination (see appendix 4). All models for inclusion appeared to originate from psychoanalytic or behavioural domains. Considerable overlap between models was observed. CT and BT shared many treatment methods, and five models that combined elements of both CT and BT were identified. IPT explicitly integrated theory and applied methods from both psychoanalytic and behavioural schools within a single manualised model. According to the way in which they divided into psychotherapeutic groups, it was possible to classify the various models into four categories.

(a) CBT

The term cognitive behavioural is now widely used in the literature and in clinical practice to describe psychotherapy interventions that share specific characteristics. Incorporating core elements of both behavioural and cognitive models, CBT challenges negative automatic thoughts and dysfunctional underlying beliefs, which mediate the relationship between negative life events and depressive symptomatology. Patients are taught to generate alternative interpretations for cognitive errors through collaborative 'hypothesis-testing' in partnership with the therapist. Patients are also required to self-monitor their target behaviours and to carry out homework tasks, including graded assignments in order to enhance skill acquisition. CBT is a manualised and time-limited intervention, usually conducted within a ten- to 20-session framework. Psychological treatments based upon cognitive theories of depression,[17] behavioural social skills approaches[16] and other conceptually and methodologically overlapping models, such as self-control therapy,[46] problem-solving therapy[47] and Lewinsohn's coping with depression course,[48] were included in this category.

* See appendix 6 for the references for each study.

(b) IPT

Based upon the work of Sullivan,[49] IPT was developed specifically as a time-limited therapy for depression by Klerman and co-workers,[50] to be used as a psychological alternative to pharmacotherapy in depression treatment trials. IPT makes no assumptions about the aetiology of depression, but uses the connection between mood and current interpersonal experiences to focus on four potential problem areas, defined as extended grief, interpersonal dispute, role transition and interpersonal deficits. The goal of IPT is to reduce depressive symptoms by improving social competence and the quality of social relationships. Treatment methods include exploration, encouragement of affect, communication analysis and behaviour change techniques. It is a manualised and time-limited intervention and is usually conducted over 16 weekly sessions.

(c) Brief PDT

Grounded in psychoanalytic principles, this group of models uses the therapeutic relationship to clarify and explore unconscious conflict. Interpretation of transference is used to facilitate the development of insight and circumscribed character change, and relief from depressive symptoms occurs with resolution of inner conflict. Brief therapy models have been developed by Malan,[51] Balint and co-workers,[52] Mann[53] and Davenloo[54] and others. Manualised brief therapy models have been developed by Luborsky[55] and Strupp and Binder,[56] and are conducted over 16–20 sessions.

(d) ST

ST is an inclusive term, often utilised in treatment outcome trials to describe an attention-placebo condition to provide a comparison to active manualised psychological interventions. Though variously named as non-directive, insight-oriented, experiential, attention-placebo, minimal contact or relaxation therapy, authors appear to most commonly locate ST within a client-centred therapy framework. Rooted in a humanistic philosophy, client-centred therapy is non-mechanistic and experiential, and utilises core conditions of empathy, acceptance and genuineness within the therapeutic relationship to facilitate self-awareness and self-determination.[57] Depressive symptomatology may be dissipated as a result of greater self-acceptance and improved self-esteem. Traditionally, client-centred therapy is undertaken within a non-manualised open-ended contract, however, an operationalised model has recently been developed.[12]

Gestalt therapy is included in the ST category because its theoretical base is humanistic in origin. In conceptualising depression as a secondary response to unrecognised anger, the focus of therapy is the magnification of the internal experience of anger through use of a series of techniques, such as two-chair dialogues. Focused-expressive therapy[58] has been developed as a manualised version of gestalt therapy. Process experiential psychotherapy, which integrates components of client-centred and gestalt therapy, is currently under development.[12]

Main hypotheses

The trials were grouped according to the table of comparisons, which was constructed to test the main hypotheses as follows.

(1) A combination of all variants of psychotherapy would be more efficacious than treatment as usual or a waiting-list control condition.
(2) CBT would be more efficacious than IPT, PDT and ST.
(3) Individual therapy would be more efficacious than group therapy.
(4) CBT would be more efficacious than treatment as usual.

List of comparisons

Twelve main comparisons were examined.

(1) All variants of psychotherapy versus treatment as usual/a waiting-list control. Where trials had two or more arms of psychotherapeutic interventions and a treatment as usual arm, the intervention arms were collapsed into one combined therapy group.
(2) CBT + CT + BT versus IPT + PDT + ST. Where trials had two arms of IPT or PDT or ST, those intervention arms were collapsed into one combined therapy group.
(3) Individual therapy versus group therapy. This comparison included trials of all variants of psychotherapy where one model only was utilised within a trial, and the comparison under examination was individual versus group intervention.
(4) CT versus BT. This comparison was designed to establish whether any significant difference in efficacy could be found between CT and BT models in order to provide support for the rationale of combining CT, BT and CBT arms into one comparison group.

(5) CBT versus treatment as usual. This comparison included CT and BT intervention arms, as well as specifically defined cognitive behavioural models. Where CT and BT arms were used within the same trial, the two groups were combined for the comparison with the treatment as usual control group. Treatment as usual included usual care/management, no treatment or a waiting-list group.
(6) CBT versus IPT. This comparison included two trials that combined IPT and PDT into one psychological model.
(7) CBT versus PDT. This comparison included two trials that combined IPT and PDT into one psychological model.
(8) CBT versus ST. ST interventions incorporated client-centred, gestalt, process-experiential, non-specific and attention-placebo therapies. Relaxation therapy was also categorised as an ST intervention despite its behavioural origins because it was explicitly utilised in trials as a placebo condition.
(9) IPT versus treatment as usual.
(10) IPT versus ST.
(11) PDT versus ST. PDT included trials that described the psychotherapy intervention as traditional insight-oriented therapy.
(12) ST versus treatment as usual.

Quantitative results

Both dichotomous and continuous outcomes and 95% CIs were calculated for each comparison using both fixed- and random-effects models as appropriate.

Three dichotomous outcomes were pooled. The main outcome, for which most data were available, was recovery at post-treatment where, immediately following treatment, trial participants were no longer deemed to have a clinically significant level of depression. Trialists generally defined recovery in a standard way as being < 10 on the BDI or < 6 on the HRSD, or other comparable measures (for example, MMPI, Raskin Global Scale for Depression (RGSD) or Symptom Checklist (SCL)-90-revised). The second outcome was whether trial participants were 'non-symptomatic' at different appropriate points of follow-up as and when follow-up dichotomous data were reported. The definition of being non-symptomatic at follow-up used by trialists was the same as for post-treatment recovery. A conservative approach was used to summarise dichotomous clinical response outcomes, simulating an intention-to-treat (ITT) analysis. All subjects initially randomised to each treatment condition were included, regardless of whether outcome data were available. A 'worst-case scenario' approach was adopted with the assumption that all dropouts were treatment failures. The final dichotomous outcome was dropouts reported at post-treatment and reflected the pooling of those trials where dropouts during the course of the trial were actually reported by group, regardless of the reason. The reporting of dropout data is recorded in appendix 11.

Where SDs were reported or supplied by authors, pooling of continuous outcomes was undertaken. Inconsistencies and omissions identified in the available continuous data suggest that the pooled analyses presented here should be interpreted with considerable caution. Four continuous outcomes were pooled. Mean change from baseline at post-treatment reflects the degree of change in symptom levels since baseline, measured using rating scales, in the experimental group compared to the comparison group. Where adequate data on mean change from baseline were reported at different points of follow-up, these were also pooled as appropriate. Mean differences at post-treatment and follow-up reflect the post-treatment and follow-up differences between the experimental and comparison groups at the different assessment timepoints. The reporting of mean change and mean differences data is recorded in appendix 11 for post-treatment and in appendix 12 for follow-up.

Where sufficient numbers of trials were pooled for individual outcomes, tests for clinical heterogeneity (severity at baseline and number of sessions) and sensitivity analyses (trial quality and the source of participant recruitment) were conducted. These subgroup analyses are reported alongside the unadjusted results, and the graphs are available from the author on request. The limitations of these analyses are presented in the discussion section.

For each comparison, the pooled estimates for different outcomes were considered in the light of findings from heterogeneity and sensitivity analyses. Data from five outcomes are also presented in the form of funnel plots to examine the potential impact of other biases, including publication bias. These are presented at the end of the pooled analyses.

Economic outcomes are included where available, although insufficient data were provided for a pooled analysis in all cases.

Comparisons between therapies

Comparison 1: all variants of psychotherapy versus treatment as usual/a waiting-list control

Post-treatment outcomes

Recovery

Thirteen trials provided sufficient data for inclusion in the pooled analysis for post-treatment recovery on a total of 886 patients (studies 16, 25, 33–35, 40, 44–46, 48, 50, 54 and 58*). The individual OR for post-treatment recovery was statistically significant in favour of psychotherapy in nine of the trials (studies 16, 25, 33–35, 44, 46, 50 and 54*) and the remaining four trials showed a trend in favour of psychotherapy over treatment as usual (studies 40, 45, 48 and 58*). The approximate χ^2 test for statistical heterogeneity was non-significant ($\chi^2 = 12.88$, degrees of freedom (df) = 12, $p = 0.38$) and, initially, a fixed-effects model was used to pool the data from these trials. The pooled OR for recovery with psychotherapy compared with treatment as usual was 3.01 (95% CI, 2.37 to 3.99) suggesting that the odds of recovery for those receiving psychotherapy were three times greater than for those receiving treatment as usual. The result was highly significant ($z = 7.66$, $p < 0.00001$; see *Figure 1* for data and graphical presentation). The pooled RR for recovery using a random-effects model was 1.84 (95% CI, 1.43 to 2.38) and was still highly significant ($z = 4.70$, $p < 0.00001$).

Tests for heterogeneity provided some additional information for interpretation.† Pooling the trials according to the degree of baseline severity resulted in a slightly reduced but still highly significant OR for recovery in favour of psychotherapy for trials in which participants were reported as having major depression (studies 25, 34, 40, 44–46, 54 and 58*). In those trials where participants' degree of baseline severity was unspecified (studies 16, 33, 48 and 50*), pooling resulted in a much increased, although less statistically significant, OR. Pooling the trials according to the number of psychotherapy sessions revealed a slight trend in the recovery data that suggested that the more sessions received the greater the degree of improvement with psychotherapy compared to with treatment as usual.

For the sensitivity analyses,† dividing the trials according to their overall quality scores revealed a pooled OR (favouring psychotherapy) for lower-quality trials (those scoring 10–19 out of 46; studies 16, 35, 50 and 58*) that was slightly lower and less precise than for high-quality trials (those scoring ≥ 20; studies 25, 33, 34, 40, 44–46, 48 and 54*), although both were still highly significant. The source of participant recruitment affected the pooled estimates considerably. A lower, although more precise, OR for the trials that recruited via outpatient clinics and referrals (studies 25, 40, 44–46, 50 and 54*) compared with the trials that recruited volunteers or responders to advertisements (studies 16, 33–35, 48 and 58*) was revealed. Again, both pooled estimates remained highly significant.

Dropouts

A total of 17 trials reported dropouts at post-treatment on a total of 774 patients (studies 6, 7, 16, 21, 30, 33–35, 38, 40, 43, 45, 46, 48, 54, 58 and 60*). The individual OR for dropouts was statistically significant in two of the trials, one in favour of psychotherapy (study 7*) and one in favour of treatment as usual (study 35*). Some trials indicated extreme differences between groups (studies 35 and 60*). One of the trials reported no dropouts in either group (study 48*). The approximate χ^2 test for heterogeneity was statistically significant ($\chi^2 = 27.09$, df = 15, $p = 0.028$) and a random-effects model was used to pool data from these trials. The pooled OR was 1.45 (95% CI, 0.81 to 2.58; $z = 1.26$, $p = 0.2$) suggesting that the odds of treatment discontinuation in those receiving psychotherapy were no greater than in those receiving treatment as usual (see *Figure 2*). The pooled RR for recovery using a random-effects method was 1.32 (95% CI, 0.83 to 2.11; $z = 1.16$, $p = 0.2$) again suggesting no difference in treatment discontinuation between the two groups.

Tests for clinical heterogeneity† suggested a decreasing trend in the size of pooled estimates from trials involving patients with severe depression (studies 34, 38, 40, 45, 46, 54 and 58*) to those involving mild/moderate depression (studies 7, 14, 21 and 35*) followed by those where baseline severity was unspecified (studies 6, 16, 30, 33, 43, 48 and 60*). However, none of the estimates were statistically significant using a random-effects model. Pooling the trials according to the number of psychotherapy sessions given also suggested no significant differences between groups in terms of dropouts, although, again, a trend was observed

* See appendix 6 for the references for each study.

† Plots derived from the heterogeneity and sensitivity analyses can be obtained from the author on request.

indicating that the number of dropouts was lowest with psychotherapy treatment in the trials offering one to six sessions (studies 7, 14, 16, 30, 46, 48 and 58*) and greatest in those offering 13 to 20 sessions (studies 6, 21, 33–35, 38, 43, 45, 60*).

In the sensitivity analyses,† dividing the trials according to their overall quality scores suggested no significant differences in dropouts between groups in both higher- and lower-quality trials, although the pooled estimate for lower-quality trials (those scoring 0–19; studies 7, 16, 21, 30, 35 and 58*) was greatest. Dividing trials according to the source of participant recruitment produced a statistically homogeneous group of trials that recruited via outpatient clinics and referrals, but for which no significant differences in dropouts were observed (studies 30, 40, 45, 46 and 54*). Similarly, pooling of trials that had recruited volunteers or responders to advertisements demonstrated no significant differences in dropouts between groups (studies 6, 7, 16, 21, 33–35, 38, 43, 48, 58 and 60*).

Mean differences

Mean differences between groups were reported by 22 trials on a total of 943 patients (studies 3, 6, 7, 9, 14, 25, 30, 33–35, 38, 40, 43–46, 48, 51, 53, 57, 58 and 60*). All but one (study 25*) of the mean differences from individual trials suggested efficacy in favour of psychotherapy, and 11 (studies 3, 7, 30, 33, 34, 35, 43, 44, 51, 53 and 60*) were statistically significant. Four of the trials reported outcomes on different scales (studies 25, 44, 45 and 51*). The approximate χ^2 test for heterogeneity was highly significant ($\chi^2 = 89.26$, df = 21, $p < 0.00001$). No single trial contributed more than 6.5% of the weight to the pooled estimate and no single trial provided an obvious source of heterogeneity. A random-effects model was used to pool the data from these trials. The SMD was –0.90 (95% CI, –1.21 to –0.60) in favour of psychotherapy, and the result was highly significant ($z = 5.77$, $p < 0.00001$), suggesting significantly fewer symptoms post-treatment in those receiving psychotherapy than in those receiving treatment as usual (see *Figure 3*).

Examination of heterogeneity† demonstrated that when trials were grouped according to the degree of baseline severity, a trend emerged that was highly significant in all categories of severity. The reduction in symptoms with psychotherapy was greatest in trials where baseline severity was unspecified (studies 6, 30, 33, 43, 48, 51 and 60*), followed by those where severity was mild/moderate (studies 3, 7, 9, 14, 35, 53 and 57*) and was lowest in those with the most severe depression (studies 25, 34, 38, 40, 44–46 and 58*). Pooling the trials according to the number of psychotherapy sessions given suggested a less marked but still highly significant effect in favour of psychotherapy for those receiving one to six sessions (studies 3, 7, 9, 14, 25, 30, 46, 48, 51, 53, 57 and 58*). A slightly more pronounced effect in favour of psychotherapy was observed in trials where between seven and 12 sessions of psychotherapy were provided (studies 6, 33–35, 38, 43, 45 and 60*).

In the sensitivity analyses,† dividing the trials according to their overall quality scores suggested a trend in which pooling trials of higher quality (scoring ≥ 20; studies 6, 25, 33, 34, 38, 40, 43–46, 48 and 60*) resulted in a more pronounced and highly significant difference in favour of psychotherapy. Pooling of trials of intermediate quality (scoring 10–19; studies 3, 9, 14, 30, 35, 53 and 58*) also demonstrated a less marked but still highly significant difference in favour of psychotherapy, whilst those of poorest quality (scoring 0–9; studies 7, 51 and 57*) demonstrated only a borderline difference. Dividing trials according to the source of recruitment suggested a small but precise and statistically significant effect in favour of psychotherapy in those that recruited via outpatient clinics and referrals (studies 25, 30, 40 and 44–46*). Pooling of trials that had recruited volunteers or responders to advertisements (studies 3, 6, 7, 9, 14, 33–35, 38, 43, 48, 51, 53, 57, 58 and 60*) demonstrated a stronger and more statistically significant difference in favour of psychotherapy.

Mean change

Six trials reported mean change in symptom levels from baseline on a total of 363 patients (studies 21, 44–46, 57 and 58*). Four of the effect sizes from individual trials were in favour of psychotherapy (studies 21, 44, 46 and 58*), although only one reached borderline significance (study 44*). This latter trial contributed over 50% of the weight to the pooled estimate. Three of the trials reported outcomes on different scales (studies 44, 45 and 57*). The approximate χ^2 test for heterogeneity was non-significant ($\chi^2 = 2.06$, df = 5, $p = 0.84$) and a fixed-effects model was used to pool the

* See appendix 6 for the references for each study.

† Plots derived from the heterogeneity and sensitivity analyses can be obtained from the author on request.

data from these trials. The SMD for mean change was 0.27 (95% CI, 0.06 to 0.48) in favour of psychotherapy. The result was statistically significant ($z = 2.54$, $p = 0.01$), suggesting a significantly greater reduction in symptoms from baseline in those receiving psychotherapy than in those receiving treatment as usual (see *Figure 4*). The small number of trials providing data for this outcome prevented any further exploration of heterogeneity.

Follow-up outcomes

Non-symptomatic at follow-up

Four trials provided sufficient data for inclusion in the pooled analysis for whether participants were non-symptomatic at follow-up on a total of 310 patients (studies 25, 35, 48 and 54*). All four reported outcomes at 3 months post-treatment. Three of these showed no significant differences in outcome between psychotherapy and treatment as usual at 3 months (studies 25, 35 and 54*). The smallest trial reported an extreme and highly significant difference in favour of psychotherapy (study 48*). The approximate χ^2 test for statistical heterogeneity was significant ($\chi^2 = 9.95$, df = 3, $p = 0.019$), and a random-effects model was used that resulted in a pooled OR of 1.59 (95% CI, 0.65 to 3.90). This result was non-significant ($z = 1.01$, $p = 0.3$) suggesting no difference in the odds of being non-symptomatic at 3 months between the two groups (see *Figure 5*). The pooled RR for recovery using a random-effects model was 1.16 (95% CI, 0.78 to 1.72; $z = 0.72$, $p = 0.5$). The small number of trials providing data for this outcome prevented any further exploration of heterogeneity.

Mean differences at follow-up

A total of 11 trials reported mean differences between groups at follow-up on a total of 594 patients (studies 3, 7, 9, 14, 25, 33, 43, 44, 46, 48 and 57*). Three of the trials reported outcomes on a different scale (studies 25, 44 and 57*). A random-effects model was used to pool the data at each timepoint. Three trials reported mean differences at up to 1 month ($n = 78$; studies 3, 9 and 14*) and two of these suggested a significant difference in favour of psychotherapy (studies 3 and 14*). Pooling suggested a statistically significant difference in favour of psychotherapy over treatment as usual (SMD = –1.61, 95% CI, –2.98 to –0.25; $z = 2.31$, $p = 0.02$). Six trials reported this outcome at up to 3 months ($n = 237$; studies 25, 33, 43, 46, 48 and 57*). All reported results in favour of psychotherapy although only three were statistically significant (studies 25, 33 and 48*). The pooled SMD suggested a robust and statistically significant difference in favour of psychotherapy over treatment as usual (SMD = –0.63, 95% CI, –1.06 to –0.20; $z = 2.87$, $p = 0.004$). Two trials reported this outcome at between 6 and 9 months ($n = 211$; studies 44 and 46*) and pooling suggested a statistically significant difference in favour of psychotherapy over treatment as usual (SMD = –0.56, 95% CI, –0.83 to –0.28; $z = 3.96$, $p = 0.00008$). One trial reported data at 1 year but was relatively small ($n = 24$) and found no significant difference (study 46*). One trial reported 2-year data ($n = 44$) and suggested a statistically significant difference in favour of psychotherapy over treatment as usual (study 7*). See *Figure 6* for data and graphical presentation. The small number of trials providing data in each category for this outcome prevented any further exploration of heterogeneity.

Mean change at follow-up

Four trials reported follow-up data on mean change in symptom levels from baseline on a total of 529 patients (studies 25, 44, 46 and 57*). All of the trials reported outcomes on different scales. Three trials reported mean change at up to 3 months ($n = 177$; studies 25, 46 and 57*) and pooling suggested a statistically significant difference in favour of psychotherapy over treatment as usual (SMD = 0.48, 95% CI, 0.17 to 0.78; $z = 3.09$, $p = 0.002$). Three trials reported this outcome at between 6 and 9 months ($n = 329$; studies 25, 44 and 46*) and pooling suggested a statistically significant difference in favour of psychotherapy over treatment as usual (SMD = 0.43 with 95% CI, 0.21 to 0.65; $z = 3.83$, $p = 0.0001$). One small trial reported this outcome at 1 year ($n = 23$) and found no difference between the two groups (study 46*). See *Figure 7* for data and graphical presentation. The small number of trials providing data in each category for this outcome prevented any further exploration of heterogeneity.

Economic outcomes

Three economic evaluations offered evidence (studies 25, 44 and 45*). None was sufficiently large to allow robust testing of efficiency differences between psychotherapy and treatment as usual, but each suggested that psychotherapy might have been superior with larger samples. The two cost-effectiveness analyses reported by

* See appendix 6 for the references for each study.

von Korff and co-workers,[59] linked to the trial of Katon and colleagues (study 25*), and Lave and colleagues[60] (a cost–utility analysis), linked to the trial by Schulberg and co-workers in 1996 (study 44*), are generally well-designed and well-conducted evaluations. They provide tentative support for the hypothesis that psychotherapy is more efficient than usual care. The UK study by Scott and Freeman in 1992 (study 45*) had smaller samples and employed a less robust economic methodology, reaching a similar conclusion.

Comparison 2: CBT + CT + BT versus PDT + IPT + ST

Post-treatment outcomes

Recovery

Sixteen trials (contributing seventeen sets of data) provided sufficient data for inclusion in the pooled analysis for post-treatment recovery on a total of 1024 patients (studies 4, 5, 11, 13, 16, 17, 26, 28, 29, 31, 33, 45, 49 (eight-session arm), 49 (16-session arm), 50, 52 and 56*). One of the trials had conducted a comparison of the number of sessions offered (eight versus 16 sessions) in addition to the comparison between models (study 49*). Data from this trial were divided into an eight-session trial and a 16-session trial and entered into the meta-analysis separately. Thirteen of the individual ORs for post-treatment recovery suggested a treatment effect in favour of variants of CBT (studies 4, 11, 16, 17, 28, 29, 33, 45, 49 (eight-session arm), 49 (16-session arm), 50, 52 and 56*), but only six of these were statistically significant (studies 11, 16, 28, 33, 52 and 56*). One trial reported a statistically significant OR in favour of IPT, PDT and ST (study 45*). The approximate χ^2 test for heterogeneity was highly significant ($\chi^2 = 58.27$, df = 16, $p < 0.00001$). A random-effects model was used to pool the data from these trials. The pooled OR for recovery with variants of CBT compared with IPT, PDT or ST was 2.40 (95% CI, 1.37 to 4.21) suggesting that the odds of recovery for those receiving variants of CBT were more than twice that of those receiving IPT, PDT and ST. The result was statistically significant ($z = 3.06$, $p = 0.002$; see *Figure 8*). The pooled RR for recovery using a random-effects model was 1.49 (95% CI, 1.11 to 2.00) and was still significant ($z = 2.63$, $p = 0.008$).

In the tests for clinical heterogeneity,† pooling the trials according to the degree of baseline severity revealed no differences in recovery between variants of CBT versus IPT, PDT and ST for patients with severe depression (studies 4, 5, 11, 13, 26, 31, 45 and 49 (both eight- and 16-session arms)*). However, a highly significant difference between groups was demonstrated by pooling those trials in which baseline severity was not specified (studies 16, 17, 28, 29, 33, 50, 52 and 56*). Pooling the trials according to the number of psychotherapy sessions revealed marked differences between trials according to the number of sessions offered. Only the pooled estimate for those trials in which one to six sessions were offered demonstrated a significant difference in favour of variants of CBT (studies 16, 52 and 56*). Although a decreasing trend in the size of the pooled estimate was observed with increasing numbers of sessions, a statistically significant difference was not demonstrated in trials offering either seven to 12 (studies 4, 26, 28, 29, 33, 45, 49 (eight-session arm) and 50*) or 13 to 20 sessions (studies 5, 11, 13, 17, 31 and 49 (16-session arm)*).

In the sensitivity analyses,† dividing the trials according to their overall quality scores suggested a trend in which pooling trials of the lowest quality (scoring 0–19; studies 16, 29, 50, 52 and 56*) resulted in a more pronounced and highly significant difference in recovery in favour of variants of CBT. Pooling of trials of intermediate quality (scoring 20–30; studies 5, 11, 17, 26, 28, 31, 33, 45 and 49 (both eight- and 16-session arms)*) demonstrated a less marked and non-significant difference in favour of variants of CBT, whilst in trials of the highest quality (scoring ≥ 30; studies 4 and 13*) pooling suggested no significant difference between groups. Dividing trials according to the source of recruitment suggested a large and highly significant difference in recovery in favour of variants of CBT only in trials that had recruited volunteers or responders to advertisements (studies 4, 5, 11, 16, 17, 26, 28, 31, 33, 52 and 56*). Pooling of trials that recruited via outpatient clinics and referrals (studies 13, 29, 45, 49 (both eight- and 16-sessions arms) and 50*) demonstrated no difference in recovery between groups.

Dropouts

Dropouts at post-treatment were reported by 19 trials (contributing 20 sets of data) on a total of 1047 patients (studies 4, 5, 11, 13, 16, 17, 26, 28–31, 33, 36, 42, 45, 49 (eight-session arm), 49 (16-session arm), 50, 52 and 56*). The individual

* See appendix 6 for the references for each study.

† Plots derived from the heterogeneity and sensitivity analyses can be obtained from the author on request.

OR for dropouts was statistically significant in two of the trials, one in favour of variants of CBT (study 28*) and one in favour of IPT, PDT and ST (study 45*). The remainder demonstrated no significant differences between the two categories of psychotherapy. The approximate χ^2 test for heterogeneity was non-significant ($\chi^2 = 23.86$, df = 18, $p = 0.16$) and a fixed-effects model was initially used to pool the data from these trials. The pooled OR was 0.91 (95% CI, 0.67 to 1.23) and the result was non-significant ($z = -0.61$, $p = 0.5$) suggesting no difference in the odds of treatment discontinuation for the two types of therapy (see *Figure 9*). The pooled RR for dropouts using a random-effects model was 0.94 (95% CI, 0.72 to 1.23; $z = -0.45$, $p = 0.7$).

Tests for clinical heterogeneity† on baseline severity revealed marked differences, with effects apparently in opposite directions. A statistically significant effect in favour of variants of CBT was observed only where baseline severity was unspecified (studies 16, 17, 28–30, 33, 42, 50, 52 and 56*). However, in trials involving patients with severe depression (studies 4, 5, 11, 13, 26, 31, 45 and 49 (both eight- and 16-session arms)*), the effect, although not quite statistically significant, appeared to demonstrate a greater dropout rate in those receiving variants of CBT. Pooling the trials according to the number of psychotherapy sessions given also suggested a trend in the size of the effect, with 13 to 20 sessions (studies 5, 11, 13, 17, 31 and 49 (16-session arm)*) demonstrating the highest number of dropouts from variants of CBT. However, the differences between groups were non-significant in all categories.

Sensitivity analyses† revealed no signs of a trend and no significant differences in dropouts between groups according to different categories of trial quality. Again, no significant differences in dropouts between groups were observed according to the source of recruitment. However, trials in which patients had been recruited through outpatient clinics and referrals (studies 13, 29, 30, 42, 45, 49 (both eight- and 16-session arms) and 50*) demonstrated greater numbers of dropouts from variants of CBT, whereas trials that had recruited volunteers or responders to advertisements (studies 4, 5, 11, 16, 17, 26, 28, 31, 33, 36, 52 and 56*) demonstrated greater numbers of dropouts from IPT, PDT and ST.

Mean differences

Mean differences between groups were reported by 13 trials on a total of 492 patients (studies 4, 5, 13, 17, 22, 26, 29–31, 33, 36, 45 and 52*). Four of the mean differences from individual trials were statistically significant in favour of variants of CBT (studies 29, 30, 33 and 52*) and one demonstrated a statistically significant difference in favour of IPT, PDT and ST (study 26*). The approximate χ^2 test for heterogeneity was highly significant ($\chi^2 = 48.65$, df = 12, $p < 0.00001$). No single trial appeared to account for the heterogeneity observed, but several trials reported very different data from the rest (studies 30, 36 and 52*), and, in addition, several extreme ORs were observed (studies 26, 30, 33 and 52*). A random-effects model was used to pool the data from these trials. The WMD was –2.05 (95% CI, –4.92 to 0.83), and the result was non-significant ($z = 1.39$, $p = 0.16$) suggesting no difference between the two types of therapy (see *Figure 10*).

Examination of heterogeneity† demonstrated that when trials were grouped according to the degree of baseline severity, the reduction in symptoms with variants of CBT was again significantly greater than with IPT, PDT and ST only in those trials in which baseline severity was unspecified (studies 17, 22, 29, 30, 33 and 52*). In trials involving patients with severe depression (studies 4, 5, 13, 26, 31 and 45*), no difference in mean scores between groups was demonstrated. Pooling the trials according to the number of psychotherapy sessions given suggested a marked difference in favour of variants of CBT in those where between one and six sessions were offered (studies 30 and 52*). Although the size of the effect appeared to reduce as the number of sessions increased, there was no significant difference demonstrated in either trials with seven to 12 sessions (studies 4, 22, 26, 29, 33, 36 and 45*) or trials with 13 to 20 sessions (studies 5, 13, 17 and 31*).

In the sensitivity analyses,† dividing the trials according to their overall quality scores suggested trials of lowest quality (scoring 0–19; studies 29, 30, 36 and 52*) resulted in a more pronounced and highly significant pooled difference in favour of variants of CBT. Pooling of trials of intermediate (scoring 20–30; studies 5, 17, 22, 26, 31, 33 and 45*) or higher (≥ 30; studies 4 and 13*) quality demonstrated no differences between groups. Dividing trials according to the source of recruit-

* See appendix 6 for the references for each study.

† Plots derived from the heterogeneity and sensitivity analyses can be obtained from the author on request.

ment demonstrated no significant differences between groups, although a more pronounced WMD in favour of variants of CBT was observed in those trials that had recruited through outpatient clinics and referrals (studies 13, 22, 29, 30, 45 and 52*).

Mean change

Five trials (contributing six data sets) reported mean change in symptom levels from baseline on a total of 283 patients (studies 5, 17, 36, 45, 49 (eight-session arm) and 49 (16-session arm)*). None of the original trials found a statistically significant difference between the two types of therapy. The approximate χ^2 test for heterogeneity was non-significant ($\chi^2 = 3.38$, df = 5, $p = 0.64$), and, initially, a fixed-effects model was used to pool the data from these trials. The WMD for mean change was 1.71 (95% CI, –0.23 to 3.65) and the result was non-significant ($z = 1.72$, $p = 0.08$), suggesting no difference in mean change from baseline between the two types of therapy (see *Figure 11*). The small number of trials providing data for this outcome prevented any further exploration of heterogeneity.

Follow-up outcomes

Non-symptomatic at follow-up

Five trials (contributing six sets of data) provided sufficient data for inclusion in the pooled analysis for whether participants were non-symptomatic at follow-up on a total of 437 patients (studies 5, 17, 28, 33, 49 (eight-session arm) and 49 (16-session arm)*; see *Figure 12*). A fixed-effects model was used to pool data at each timepoint. Three trials reported outcomes at 3 months post-treatment ($n = 264$; studies 5, 17 and 28*) and pooling demonstrated no significant difference between variants of CBT and IPT, PDT and ST (OR = 1.49, 95% CI, 0.87 to 2.54; $z = 1.44$, $p = 0.15$). One small trial reported outcomes at 6 months post-treatment ($n = 23$; study 33*) and had an extremely high OR in favour of variants of CBT. Two sets of data were available for outcomes at 1 year, which were two arms given different numbers of sessions within the same trial (study 49*; total $n = 160$). Pooling of these two arms demonstrated no significant difference between variants of CBT and IPT, PDT and ST. The small number of trials providing data for this outcome prevented any further exploration of heterogeneity.

Mean differences at follow-up

A total of six trials reported mean differences between groups at follow up on a total of 170 patients (studies 5, 17, 22, 26, 29 and 33*; see *Figure 13*). A fixed-effects model was used to pool the data at each timepoint. Five trials reported mean differences at up to 3 months ($n = 152$; studies 5, 17, 22, 26 and 29*) and pooling suggested a statistically significant difference in favour of CBT compared to IPT, PDT and ST (WMD = –2.48, 95% CI, –4.26 to –0.71; $z = 2.74$, $p = 0.006$). One small trial (study 33*) reported this outcome at 6 months ($n = 18$) and found a statistically significant difference in favour of variants of CBT. The small number of trials providing data for this outcome prevented any further exploration of heterogeneity.

Mean change at follow-up

Two trials reported data on mean change in symptom levels from baseline to a follow-up of 3 months on a total of 91 patients (studies 5 and 17*; see *Figure 14*). Neither trial found any difference between the two categories of therapy, and the overall pooled estimate using a fixed-effects model also demonstrated no differences (WMD = 1.16, 95% CI, –2.27 to 4.59; $z = 0.66$, $p = 0.5$). The small number of trials providing data for this outcome prevented any further exploration of heterogeneity.

Economic outcomes

Only one trial included an economic component (study 46*), but there were insufficient data for analysis.

Comparison 3: individual versus group therapy

Post-treatment outcomes

Recovery

Six trials provided sufficient data for inclusion in the pooled analysis for post-treatment recovery on a total of 231 patients (studies 40, 41, 47, 55, 59 and 63*). The trials all utilised cognitive or behavioural interventions. The individual OR for post-treatment recovery was statistically significant in favour of individual therapy in only one trial (study 59*), with the remaining trials demonstrating no significant differences between individual and group treatments. Statistical heterogeneity was non-significant ($\chi^2 = 4.94$, df = 5, $p = 0.42$) and a fixed-effects method was used. The pooled OR for recovery with individual compared to group therapy was 1.98 (95% CI, 1.11 to 3.54) suggesting that the odds of recovery for those receiving individual therapy were nearly twice that of those receiving

* See appendix 6 for the references for each study.

group therapy. The result was statistically significant ($z = 2.32$, $p = 0.02$; see *Figure 15*). The pooled RR for recovery using a random-effects model was 1.24 (95% CI, 1.01 to 1.53). The small number of trials providing data for this outcome prevented any further exploration of heterogeneity.

Dropouts

Six trials reported dropouts at post-treatment on a total of 217 patients (studies 6, 40, 41, 43, 47 and 63*). The approximate χ^2 test for heterogeneity was non-significant ($\chi^2 = 3.97$, df = 5, $p = 0.55$) and, therefore, a fixed-effects model was used to pool the data from these trials. The pooled OR was 0.47 (95% CI, 0.21 to 1.09) and the result was non-significant ($z = -1.76$, $p = 0.08$) suggesting no differences in treatment discontinuation between individual and group therapy (see *Figure 16*). The pooled RR for dropouts using a random-effects model was 0.57 (95% CI, 0.32 to 1.04). The small number of trials providing data for this outcome prevented any further exploration of heterogeneity.

Mean differences

Eight trials reported mean differences between groups on a total of 283 patients (studies 6, 40, 41, 43, 47, 55, 59 and 63*). Most results were in favour of individual treatment, although only two of the mean differences from individual trials (studies 41 and 55*) were statistically significant. One of these contributed nearly 60% of the weight to the pooled estimate (study 55*). The approximate χ^2 test for heterogeneity was non-significant ($\chi^2 = 6.71$, df = 7, $p = 0.46$) and a fixed-effects model was used to pool the data from these trials. The WMD was –3.07 (95% CI, –4.69 to –1.45) in favour of individual treatment and the result was highly significant ($z = 3.71$, $p = 0.0002$), suggesting that reduction in symptoms was significantly greater in those receiving individual therapy than in those receiving therapy in a group. See *Figure 17* for data and graphical presentation. The small number of trials providing data for this outcome prevented any further exploration of heterogeneity.

Mean change

No trials reported data on mean change in symptom levels from baseline for this comparison.

Follow-up outcomes

Non-symptomatic at follow-up

Three trials provided sufficient data for inclusion in the pooled analysis for whether participants were non-symptomatic at follow-up on a total of 143 patients (studies 6, 55 and 63*; see *Figure 18*). A fixed-effects model was used to pool data at each timepoint. One small trial reported outcomes at 2 months post-treatment ($n = 30$; study 63*) and pooling demonstrated no difference between individual and group therapy. Two trials (studies 6 and 55*) reported outcomes at 6 months post-treatment ($n = 113$). Both the individual ORs and the pooled OR demonstrated no differences between individual and group treatments (OR = 1.28, 95% CI, 0.52 to 3.20; $z = 0.54$, $p = 0.6$). The small number of trials providing data for this outcome prevented any further exploration of heterogeneity.

Mean differences at follow-up

A total of six trials reported mean differences between groups at follow-up on a total of 407 patients (studies 6, 40, 43, 47, 55 and 63*; see *Figure 19*). A random-effects model was used to pool the data at each timepoint. Two trials reported mean differences at 1 month ($n = 74$), one demonstrating a statistically significant difference in favour of group therapy (study 6*) and one demonstrating no difference (study 47*). The overall pooled estimate also demonstrated no statistically significant difference between the two formats (WMD = 5.20, 95% CI, 0.21 to 10.19; $z = 2.04$, $p = 0.04$). Three trials reported this outcome at 2 months ($n = 84$; studies 43, 47 and 63*) and all demonstrated no differences between the individual and group formats, either individually or when pooled (WMD = 0.21, 95% CI, –3.53 to 3.96; $z = 0.11$, $p = 0.9$). Two trials reported mean differences data at a follow-up of 3 months ($n = 65$; studies 40 and 47*) and, again, demonstrated no differences either individually or once pooled (WMD = 0.25, 95% CI, –4.61 to 5.11; $z = 0.10$, $p = 0.9$). Four trials reported 6-month data ($n = 155$), one of which demonstrated a significant difference in favour of group therapy (study 40*) whereas the remainder reported no differences (studies 6, 47 and 55*). Pooling of these trials also demonstrated no difference between the two formats (WMD = 3.21, 95% CI, –2.18 to 8.60; $z = 1.17$, $p = 0.2$). One trial reported mean differences at 1-year follow-up ($n = 29$) and suggested a significant difference again in favour of group therapy (study 40*). The small number of trials providing data for this outcome prevented any further exploration of heterogeneity.

* See appendix 6 for the references for each study.

Mean change
No trials reported data on mean change in symptom levels from baseline for this comparison.

Economic outcomes
There was no economic evidence for this comparison.

Comparison 4: CT versus BT
Post-treatment outcomes
Recovery
Three trials provided sufficient data for inclusion in the pooled analysis for post-treatment recovery on a total of 149 patients (studies 23, 29 and 50*). None of the individual trials demonstrated a significant difference between CT and BT. Statistical heterogeneity was non-significant (χ^2 = 1.34, df = 2, p = 0.51) and, therefore, a fixed-effects model was used to pool the data from these trials. The pooled OR was 1.58 (95% CI, 0.83 to 2.99) and was non-significant (z = 1.39, p = 0.16) suggesting no difference between CT and BT (see *Figure 20*). The pooled RR for recovery using a random-effects model was 1.21 (95% CI, 0.89 to 1.65). The small number of trials providing data for this outcome prevented any further exploration of heterogeneity.

Dropouts
Three trials reported dropouts at post-treatment (n = 149; studies 23, 29 and 60*). None of the individual trials demonstrated a difference between CT and BT. The approximate χ^2 test for heterogeneity was non-significant (χ^2 = 2.73, df = 2, p = 0.25) and a fixed-effects model was thus used to pool the data from these trials. The pooled OR was 0.83 (95% CI, 0.30 to 2.24) and was, therefore, non-significant (z = 0.38, p = 0.7) suggesting no difference in treatment discontinuation between CT and BT (see *Figure 21*). The small number of trials providing data for this outcome prevented any further exploration of heterogeneity.

Mean differences
Six trials reported mean differences between groups on a total of 247 patients (studies 23, 29, 39, 53, 60 and 62*). None of the mean differences from individual trials demonstrated a difference between CT and BT. The approximate χ^2 test for heterogeneity was non-significant (χ^2 = 2.12, df = 5, p = 0.83) and a fixed-effects model was used to pool the data from these trials. The WMD was 0.61 (95% CI, –1.08 to 2.30) and was, therefore, non-significant (z = 0.71, p = 0.5) suggesting no difference between CT and BT (see *Figure 22*). The small number of trials providing data for this outcome prevented any further exploration of heterogeneity.

Mean change
No trials reported data on the mean change in symptom levels from baseline for this comparison.

Follow-up outcomes
Non-symptomatic at follow-up
Only one trial (study 23*) reported data on whether participants were non-symptomatic at follow-up (n = 107) and demonstrated no difference between CT and BT.

Mean differences at follow-up
A total of six trials reported mean differences between groups at follow-up on a total of 217 patients (studies 23, 29, 39, 53, 60 and 63*; see *Figure 23*). A fixed-effects model was used to pool the data at each timepoint. Of these, three trials reported mean differences at 2 months (n = 55; studies 29, 53 and 63*), all demonstrating no differences between CT and BT, both individually and once pooled (WMD = –1.20, 95% CI, –2.81 to 0.42; z = 1.45, p = 0.15). The other three trials (studies 23, 39 and 60*) reported this outcome at between 2 and 6 months (n = 162) and, again, all demonstrated no differences between CT and BT either individually or once pooled (WMD = 1.24, 95% CI, –1.06 to 3.55; z = 1.06, p = 0.3). The small number of trials providing data for this outcome prevented any further exploration of heterogeneity.

Mean change at follow-up
No trials reported data on mean change in symptom levels from baseline to follow-up for this comparison.

Economic outcomes
There was no economic evidence available for this comparison.

Comparison 5: CBT + CT + BT versus treatment as usual/a waiting-list control
Post-treatment outcomes
Recovery
Sufficient data were provided by 12 trials for inclusion in the pooled analysis for post-treatment recovery on a total of 654 patients (studies 16, 25, 33–35, 40, 45, 46, 48, 50, 54 and 58*). The

* See appendix 6 for the references for each study.

individual OR for post-treatment recovery was statistically significant in favour of variants of CBT in seven of the trials (studies 16, 25, 33–35, 46 and 54*) and the remaining five demonstrated no significant differences from treatment as usual (studies 40, 45, 48, 50 and 58*). The approximate χ^2 test for heterogeneity was statistically significant ($\chi^2 = 21.20$, df = 11, $p = 0.031$) and a random-effects model was used to pool the data from these trials. The pooled OR for recovery with variants of CBT compared with treatment as usual was 3.42 (95% CI, 1.98 to 5.93) suggesting that the odds of recovery for those receiving variants of CBT were more than three times greater than for those receiving usual treatment. The result was highly significant ($z = 4.39$, $p = 0.00001$; see *Figure 24*). The pooled RR for recovery using a random-effects model was 1.81 (95% CI, 1.32 to 2.48).

Tests for heterogeneity provided some additional information for interpretation.† Pooling the trials according to the degree of baseline severity resulted in a slightly reduced but still highly significant pooled estimate for recovery in favour of variants of CBT for trials in which participants had more severe depression (studies 25, 34, 40, 45, 46, 54 and 58*). In those trials in which the degree of baseline severity was unspecified (studies 16, 33, 48 and 50*), pooling resulted in an even more pronounced but much less precise pooled estimate in favour of CBT. However, pooling the trials according to the number of psychotherapy sessions demonstrated few differences between groups.

For the sensitivity analyses,† dividing the trials according to their overall quality scores revealed highly significant differences in all categories favouring variants of CBT. An apparent trend suggested that the greatest effect occurred in the lowest-quality trials (scoring 10–19; studies 16, 35, 50 and 58*). Once more, dividing the trials according to the source of recruitment revealed a lower, although more precise, OR for those recruited via outpatient clinics and referrals (studies 25, 40, 45, 46, 50 and 54*) compared with trials that recruited volunteers or responders to advertisements (studies 16, 33–35, 48 and 58*). Again, both pooled estimates remained highly significant.

Dropouts

Dropouts at post-treatment for this comparison were reported by 18 trials on a total of 760 patients (studies 6, 7, 14, 16, 21, 30, 33–35, 38, 40, 43, 45, 46, 48, 54, 58 and 60*). All but three (studies 7, 16 and 35*) of the individual ORs for dropouts suggested there were no significant differences between variants of CBT and treatment as usual. The approximate χ^2 test for heterogeneity was statistically significant ($\chi^2 = 29.48$, df = 16, $p = 0.021$) and a random-effects model was used to pool the data from these trials. The pooled OR was 1.67 (95% CI, 0.92 to 3.01) and was non-significant ($z = 1.69$, $p = 0.09$) suggesting no significant difference in the odds of treatment discontinuation in those receiving variants of CBT compared with those receiving treatment as usual (see *Figure 25*). The pooled RR for recovery using a random-effects model was 1.49 (95% CI, 0.92 to 2.40).

Tests for clinical heterogeneity† revealed significantly greater numbers of dropouts from variants of CBT in trials involving more severely depressed patients (studies 34, 38, 40, 45, 46, 54 and 58*). No significant differences in dropouts between groups were observed in those trials involving mild/moderate participants (studies 7, 14, 21 and 35*) or where severity was unspecified (studies 6, 16, 30, 33, 43, 48 and 60*). Pooling the trials according to the number of psychotherapy sessions given revealed no difference in dropouts between groups when between one and six sessions were offered (studies 7, 14, 16, 30, 46, 48 and 58*). Statistically significant differences between groups, favouring treatment as usual, were, however, observed in trials where seven to 12 (studies 6, 21, 33–35, 38, 43, 45 and 60*) and 13 to 20 (studies 40 and 54*) sessions were offered.

In the sensitivity analyses,† dividing the trials according to their overall quality scores suggested a significant difference in dropouts between groups in the trials of lower quality only (studies 7, 14, 16, 21, 30, 35 and 58*). In higher-quality trials (studies 6, 33, 34, 38, 40, 43, 45, 46, 48, 54 and 60*), the difference had only borderline significance. Dividing trials according to the source of participant recruitment produced a statistically homogeneous group of trials that recruited from outpatient clinics and referrals (studies 30, 40, 45, 46 and 54*), but for which the difference in dropouts had only borderline significance in favour of treatment as usual. Pooling of trials that had recruited volunteers or responders to advertisements (studies 6, 7, 14, 16, 21, 33–35, 38, 43, 48,

* See appendix 6 for the references for each study.

† Plots derived from the heterogeneity and sensitivity analyses can be obtained from the author on request.

58 and 60*) produced a heterogeneous group of trials, which, again, only reached borderline significance in favour of treatment as usual.

Mean differences

Mean differences between groups were reported by 20 trials on a total of 748 patients (studies 3, 6, 7, 9, 14, 25, 30, 33–35, 38, 40, 43, 45, 46, 48, 51, 53, 58 and 60*). Ten of the mean differences from individual trials showed no statistically significant differences between variants of CBT and treatment as usual (studies 6, 9, 14, 25, 38, 40, 45, 46, 48 and 58*). The approximate χ^2 test for heterogeneity was highly significant ($\chi^2 = 87.08$, df = 19, $p < 0.00001$) and a random-effects model was used to pool the data from these trials. The SMD was –1.0 (95% CI, –1.35 to –0.64). The result was highly significant ($z = 5.51$, $p < 0.00001$) suggesting that reduction in symptoms was significantly greater in those receiving variants of CBT than in those receiving treatment as usual (see *Figure 26*).

Examination of heterogeneity† demonstrated an apparent trend of a decreasing size of effect with increasing severity. All pooled estimates of mean differences were in favour of variants of CBT and all were statistically significant, although statistical heterogeneity between trial results remained significant in all categories. Pooling the trials according to the number of psychotherapy sessions given did not reduce the observed statistical heterogeneity between trials and revealed little difference in the size of the pooled estimates, although all remained statistically significant.

In the sensitivity analyses,† dividing the trials according to their overall quality scores suggested a trend in which trials of higher quality (scoring ≥ 20; studies 6, 25, 33, 34, 38, 40, 43, 45, 46, 48 and 60*) resulted in a less pronounced effect in favour of variants of CBT, and the lowest scoring trials (scoring 0–9; studies 7 and 51*) resulted in the largest effect in favour of variants of CBT. Differences between groups were significant in all quality categories, although statistical heterogeneity remained. Investigating the role of the source of recruitment suggested a difference of only borderline significance between groups in those trials that recruited via outpatient clinics and referrals (studies 25, 30, 40, 45 and 46*). However, in those trials that had recruited volunteers or responders to advertisements (studies 3, 6, 7, 9, 14, 33–35, 38, 43, 48, 51, 53, 58 and 60*), a marked and highly significant difference in favour of variants of CBT was observed. In both categories, statistical heterogeneity remained.

Mean change

Five trials reported mean change in symptom levels from baseline on a total of 172 patients (studies 19, 21, 45, 46 and 58*). None of these trials found statistically significant differences between the two conditions. The approximate χ^2 test for heterogeneity was non-significant ($\chi^2 = 2.24$, df = 4, $p = 0.69$) and a fixed-effects model was used to pool the data from these trials. The WMD for mean change was 2.38 (95% CI, 0.05 to 4.71), which reached borderline significance ($z = 2.00$, $p = 0.05$), suggesting a possible difference in mean change from baseline in favour of variants of CBT compared with treatment as usual (see *Figure 27*). The small number of trials providing data for this outcome prevented any further exploration of heterogeneity.

Follow-up outcomes

Non-symptomatic at follow-up

Four trials provided sufficient data for inclusion in the pooled analysis for whether participants were non-symptomatic at 3-month follow-up on a total of 310 patients (studies 25, 35, 48 and 54*; see *Figure 28*). A random-effects model was used to pool data for this outcome. One small trial reported a statistically significant and extremely high OR in favour of variants of CBT (study 48*), whilst the remainder demonstrated no differences. The pooled OR also suggested that there was no difference between variants of CBT and treatment as usual (OR = 1.59, 95% CI, 0.65 to 3.90; $z = 1.01$, $p = 0.3$). The pooled RR using a random-effects model was 1.16 (95% CI, 0.78 to 1.72). The small number of trials providing data for this outcome prevented any further exploration of heterogeneity.

Mean differences at follow-up

A total of ten trials reported mean differences between groups at follow-up on a total of 579 patients (studies 3, 7, 9, 14, 25, 33, 35, 43, 46 and 48*; see *Figure 29*). A random-effects model was used to pool the data at each timepoint. Four trials reported mean differences at up to 1 month ($n = 152$; studies 3, 9, 14 and 35*) of which two demonstrated statistically significant differences in favour of variants of CBT (studies 3 and 14*).

* See appendix 6 for the references for each study.

† Plots derived from the heterogeneity and sensitivity analyses can be obtained from the author on request.

The pooled estimate demonstrated a statistically significant difference in favour of variants of CBT (SMD = –1.20, 95% CI, –2.10 to –0.31; $z = 2.63$, $p = 0.009$). Two trials reported this outcome at 2 months ($n = 44$; studies 43 and 48*) and pooling these suggested no significant difference between variants of CBT and treatment as usual (SMD = –0.73, 95% CI, –1.80 to 0.33; $z = 1.35$, $p = 0.18$). Two trials reported this outcome at 3 months ($n = 155$; studies 25 and 46*) and pooling of these suggested a significant difference in favour of variants of CBT (SMD = –0.36, 95% CI, –0.68 to –0.04; $z = 2.19$, $p = 0.03$). Three trials reported this outcome at 6 months ($n = 160$), one of which demonstrated a significant difference in favour of variants of CBT (study 33*) and the remainder reported no differences (studies 25 and 46*). Pooling of these trials resulted in no significant difference between variants of CBT and treatment as usual (SMD = –0.73, 95% CI, –1.50 to 0.03; $z = 1.89$, $p = 0.06$). One trial reported mean differences data at a follow-up of 1 year ($n = 24$) and suggested no significant difference between the two conditions (study 46*). One trial reported mean differences data at 2-year follow-up ($n = 44$; study 7*) and suggested a statistically significant difference in favour of variants of CBT. The small number of trials providing data for this outcome prevented any further exploration of heterogeneity.

Mean change at follow-up

Two trials reported follow-up data on mean change in symptom levels from baseline (studies 25 and 46*; see *Figure 30*). Both reported mean change at 3 months ($n = 155$) and pooling suggested a statistically significant difference in favour of variants of CBT over treatment as usual (WMD = 7.11, 95% CI, 3.25 to 10.98; $z = 3.61$, $p = 0.0003$). Both also reported this outcome at 6 months ($n = 144$) and, again, pooling suggested a statistically significant difference in favour of variants of CBT over treatment as usual (WMD = 6.41, 95% CI, 2.48 to 10.34; $z = 3.20$, $p = 0.001$). Only one of the trials reported this outcome at 1 year ($n = 23$), and found no significant difference between the two groups (study 46*). The small number of trials providing data for this outcome prevented any further exploration of heterogeneity.

Economic outcomes

No cost-effectiveness analyses were available for any of the included trials in this comparison. One other cost-effectiveness analysis linked to an excluded trial by Katon (1995; see appendix 7 for the reference for this study), concluded that CBT offered within a collaborative care model for major depression was more costly than treatment as usual, but achieved greater success. This translated into a modest cost-effectiveness advantage. The same intervention for minor depression was more costly, but not more cost-effective than treatment as usual.

Comparison 6: CBT + CT + BT versus IPT

Post-treatment outcomes

Recovery

Two trials (contributing three data sets) provided sufficient data for inclusion in the pooled analysis for post-treatment recovery on a total of 275 patients (studies 13, 49 (eight-session arm) and 49 (16-session arm)*; see *Figure 31*). None of the individual trials demonstrated any differences in recovery between variants of CBT and IPT. The approximate χ^2 test for heterogeneity was non-significant ($\chi^2 = 2.24$, df = 2, $p = 0.33$). Applying a fixed-effects model resulted in a pooled OR of 1.08 (95% CI, 0.67 to 1.73) that was not significant ($z = 0.30$, $p = 0.8$), demonstrating no difference between the two types of therapy. The pooled RR for recovery using a random-effects model was 1.03 (95% CI, 0.78 to 1.37). The small number of trials providing data for this outcome prevented any further exploration of heterogeneity.

Dropouts

The same three trials reported dropouts at post-treatment on 275 patients (see *Figure 32*). None of the individual trials demonstrated any differences between variants of CBT and IPT. The approximate χ^2 test for heterogeneity was non-significant ($\chi^2 = 0.38$, df = 2, $p = 0.83$), and a fixed-effects model was used to pool the data from these trials. The pooled OR was 1.46 (95% CI, 0.84 to 2.56) and was non-significant ($z = 1.33$, $p = 0.18$) suggesting no difference in treatment discontinuation between variants of CBT and IPT. The pooled RR for dropouts using a random-effects model was 1.34 (95% CI, 0.88 to 2.05). The small number of trials providing data for this outcome prevented any further exploration of heterogeneity.

Mean differences

Only one trial reported mean differences between groups for this comparison on a total of 120 patients (study 13*), and demonstrated

* See appendix 6 for the references for each study.

no difference between the two types of therapy (OR = 1.40, 95% CI, –2.39 to 5.19; $z = 0.72$, $p = 0.5$).

Mean change
One trial reported data on mean change in symptom levels from baseline for this comparison (study 49*). This trial conducted a comparison of the number of sessions offered (eight versus 16) in addition to the comparison between models (total $n = 113$) and, again, demonstrated no difference between the two types of therapy (WMD = 2.80, 95% CI, –0.08 to 5.68; $z = 1.90$, $p = 0.06$).

Follow-up outcomes

Non-symptomatic at follow-up
One trial reported data for whether participants were non-symptomatic at 1-year follow-up (study 49*). This trial conducted a comparison of the number of sessions offered (eight versus 16) in addition to the comparison between models (total $n = 150$) and demonstrated no difference between the two types of therapy (OR = 1.06, 95% CI, 0.54 to 2.06; $z = 0.17$, $p = 0.8$).

Mean differences at follow-up
No trials reported data on mean differences between groups at follow-up for this comparison.

Mean change at follow-up
No trials reported data on mean change in symptom levels from baseline at follow-up.

Economic outcomes

There was no economic evidence available for this comparison.

Comparison 7: CBT + CT + BT versus PDT

Post-treatment outcomes

Recovery
Six trials (contributing seven data sets) provided sufficient data for inclusion in the pooled analysis for post-treatment recovery on a total of 484 patients (studies 4, 11, 17, 26, 28, 49 (eight-session arm) and 49 (16-session arm)*). The individual OR for post-treatment recovery was statistically significant in favour of variants of CBT in two of the trials (studies 11 and 28*) whilst the remainder demonstrated no significant differences between the two types of therapy. One of the trials (study 11*) produced extremely high positive ORs. The approximate χ^2 test for heterogeneity reached borderline significance ($\chi^2 = 12.76$, df = 6, $p = 0.047$) and a random-effects model was used to pool the data from these trials. The pooled OR was 2.11 (95% CI, 1.17 to 3.81) suggesting that the odds of recovery for those receiving variants of CBT were more than two times greater than for those receiving PDT. The result was statistically significant ($z = 2.49$, $p = 0.01$; see *Figure 33*). The pooled RR for recovery using a random-effects model was 1.43 (95% CI, 0.91 to 2.23). The small number of trials providing data for this outcome prevented any further exploration of heterogeneity.

Dropouts
Seven trials (contributing eight data sets) reported dropouts at post-treatment on a total of 512 patients (studies 4, 11, 17, 26, 28, 42, 49 (eight-session arm) and 49 (16-session arm)*). Only one of the individual trials demonstrated a statistically significant difference in the number of dropouts in favour of variants of CBT (study 28*) whilst the remainder demonstrated no differences between the two types of therapy. The approximate χ^2 test for heterogeneity was non-significant ($\chi^2 = 7.35$, df = 7, $p = 0.39$) and a fixed-effects model was used to pool the data from these trials. The pooled OR was 0.73 (95% CI, 0.47 to 1.15) and was non-significant ($z = –1.35$, $p = 0.18$) suggesting no differences in treatment discontinuation between variants of CBT and PDT (see *Figure 34*). The pooled RR for dropouts using a random-effects model was 0.83 (95% CI, 0.57 to 1.21). The small number of trials providing data for this outcome prevented any further exploration of heterogeneity.

Mean differences
Three trials reported mean differences between groups on a total of 114 patients (studies 4, 17 and 26*). Two of the mean differences from individual trials showed no statistically significant differences between variants of CBT and PDT (studies 4 and 17*), whereas the other trial reported an extreme and highly significant result in favour of PDT (study 26*). The approximate χ^2 test for heterogeneity was statistically significant ($\chi^2 = 16.05$, df = 2, $p = 0.0003$) and a random-effects model was used to pool the data from these trials. The WMD was 2.18 (95% CI, –5.90 to 10.26) and was non-significant ($z = 0.53$, $p = 0.6$) suggesting that there was no difference between the two types of therapy (see *Figure 35*). The small number of trials providing data for this outcome prevented any further exploration of heterogeneity.

* See appendix 6 for the references for each study.

Mean change
Two trials (contributing three sets of data) reported mean changes in symptom levels from baseline in a total of 165 patients (studies 17, 49 (eight-session arm) and 49 (16-session arm)[*]). None of the trials demonstrated a statistically significant difference between groups either individually or once pooled (WMD = 1.99, 95% CI, –0.49 to 4.47; $z = 1.57$, $p = 0.12$; see *Figure 36*). The small number of trials providing data for this outcome prevented any further exploration of heterogeneity.

Follow-up outcomes

Non-symptomatic at follow-up
A total of three trials (contributing four sets of data) reported data for inclusion in the pooled analysis for whether participants were non-symptomatic at follow-up (studies 17, 28, 49 (eight-session arm) and 49 (16-session arm)[*]; see *Figure 37*). Data were pooled using a fixed-effects model. Two trials reported 3-month follow-up ($n = 161$; studies 17 and 28[*]) and demonstrated no differences between the two types of therapy, either individually or once pooled (OR = 1.86, 95% CI, 0.97 to 3.58; $z = 1.86$, $p = 0.06$). One trial reported follow-up for this outcome at 1 year (study 49[*]). This trial conducted a comparison of the number of sessions offered (eight versus 16) in addition to the comparison between models, and thus contributed two sets of data (total $n = 150$). This trial also demonstrated no differences between the two types of therapy at follow-up (OR = 1.33, 95% CI, 0.69 to 2.59; $z = 0.85$, $p = 0.4$). The small number of trials providing data for this outcome prevented any further exploration of heterogeneity.

Mean differences at follow-up
Two trials (studies 17 and 26[*]) reported mean differences between groups at 3-month follow-up ($n = 63$; see *Figure 38*). Both trials suggested differences in opposite directions, but neither was significant. Using a fixed-effects model the pooled estimate demonstrated no difference between the two therapy types at follow-up (WMD = –1.39, 95% CI, –5.22 to 2.43; $z = 0.71$, $p = 0.5$). The small number of trials providing data for this outcome prevented any further exploration of heterogeneity.

Mean change at follow-up
One trial reported mean change in symptom levels from baseline at 3-month follow-up in a total of 48 patients (study 17[*]), and suggested a significant difference in favour of variants of CBT (WMD = 1.45, 95% CI, –2.99 to 5.89; $z = 0.64$, $p = 0.5$).

Economic outcomes

There was no economic evidence available for this comparison.

Comparison 8: CBT + CT + BT versus ST

Post-treatment outcomes

Recovery
Ten trials provided sufficient data for inclusion in the pooled analysis for post-treatment recovery on a total of 409 patients (studies 5, 16, 28, 29, 31, 33, 45, 50, 52 and 56[*]). The individual ORs for post-treatment recovery were statistically significant in favour of variants of CBT in five trials (studies 16, 28, 33, 52 and 56[*]), and one of the remaining trials demonstrated a statistically significant difference in favour of ST (study 45[*]). This trial had been excluded in two other comparisons on the basis that it appeared to be a CBT intervention. Two of the trials reported statistically significant and very high positive ORs (studies 33 and 52[*]). Statistical heterogeneity was highly significant ($\chi^2 = 39.98$, df = 9, $p < 0.00001$) and a random-effects method was used. The pooled OR for recovery was 3.45 (95% CI, 1.30 to 9.17) suggesting that the odds of recovery for those receiving variants of CBT was nearly three and a half times that of those receiving ST. The result was statistically significant ($z = 2.48$, $p = 0.01$; see *Figure 39*). The pooled RR for recovery using a random-effects model was 1.76 (95% CI, 1.09 to 2.84).

Tests for heterogeneity[†] provided some additional information for interpretation. Although the difference between groups was not statistically significant, the pooled estimate appeared to favour ST in those trials involving the most severely depressed patients (studies 5, 31 and 45[*]) whilst those trials in which severity was unspecified (studies 16, 28, 29, 33, 50, 52 and 56[*]) demonstrated a statistically significant difference in favour of variants of CBT. In both cases, statistical heterogeneity was removed. Pooling the trials according to the number of psychotherapy sessions suggested a decreasing trend in effect size (favouring variants of CBT) with increasing numbers of sessions. Trials that offered 13–20 sessions (studies 5 and 31[*]) did not demonstrate a significant difference between groups.

[*] See appendix 6 for the references for each study.

[†] Plots derived from the heterogeneity and sensitivity analyses can be obtained from the author on request.

In the sensitivity analyses,† dividing the trials according to their overall quality scores revealed a pooled OR favouring variants of CBT for the statistically homogeneous lower-quality trials (scoring 0–19; studies 16, 29, 50, 52 and 56*) only. In trials scoring 20–29 (studies 5, 28, 31, 33 and 45*), no significant difference in recovery was apparent, although statistical heterogeneity remained. In addition, in those trials that had recruited via outpatient clinics and referrals (studies 29, 45 and 50*), no difference between groups was demonstrated, whereas in those that had recruited volunteers or responders to advertisements (studies 5, 16, 28, 31, 33, 52 and 56*), a significant difference in favour of variants of CBT was observed. However, statistical heterogeneity remained in both groups.

Dropouts

Nine trials reported dropouts at post-treatment on a total of 377 patients (studies 16, 28–31, 33, 36, 45 and 52*). One of the individual trials demonstrated a statistically significant difference in the number of dropouts in favour of variants of CBT (study 30*) and one found in favour of ST (study 45*), whilst the remainder demonstrated no differences between the two types of therapy. The approximate χ^2 test for heterogeneity was non-significant ($\chi^2 = 14.42$, df = 8, $p = 0.071$) and a fixed-effects model was used to pool the data from these trials. The pooled OR was 0.80 (95% CI, 0.47 to 1.39) and the result was non-significant ($z = -0.78$, $p = 0.4$) suggesting no difference in treatment discontinuation between variants of CBT and ST (see *Figure 40*). The pooled RR for dropouts using a random-effects model was 0.81 (95% CI, 0.46 to 1.43).

In trials involving patients with severe depression (studies 31 and 45*), a borderline significant difference in favour of ST was observed, whilst in those where baseline severity was unspecified (studies 16, 28–30, 33 and 52*), a non-significant difference in favour of variants of CBT was suggested. Pooling the trials according to the number of psychotherapy sessions made little difference to the overall conclusion.†

Although the differences were non-significant in both cases, fewer dropouts from variants of CBT were indicated in lower-quality trials (scoring 0–19; studies 16, 29, 30, 36 and 52*) and fewer from ST were indicated for higher-quality trials (scoring 20–29; studies 28, 31, 33 and 45*). No differences in dropout rates between groups were suggested according to the source of participant recruitment.†

Mean differences

Nine trials reported mean differences between groups on a total of 261 patients (studies 5, 22, 29–31, 33, 36, 45 and 52*). Four of the mean differences from individual trials were in favour of variants of CBT (studies 29, 30, 33 and 52*) whilst the remainder demonstrated no differences (studies 5, 22, 31, 36 and 45*). The approximate χ^2 test for heterogeneity was highly significant ($\chi^2 = 29.19$, df = 8, $p = 0.0003$) and a random-effects model was used to pool the data from these trials. The WMD was –3.86 (95% CI, –7.38 to –0.33) in favour of variants of CBT, and the result was statistically significant ($z = 2.14$, $p = 0.03$) suggesting that the reduction in symptoms was greater in those receiving variants of CBT than in those receiving ST (see *Figure 41*).

Although, overall, the difference was statistically significant, dividing the trials according to the degree of baseline severity demonstrated a significant difference between groups only in the trials in which baseline severity was unspecified (studies 22, 29, 30, 33 and 52*), although statistical heterogeneity still remained in this group of trials. In the more statistically homogeneous group of trials involving severely depressed patients (studies 5, 31 and 45*) and those with mild/moderate depression (study 36*), no differences between groups were demonstrated. Pooling the trials according to the number of psychotherapy sessions given demonstrated a marked and highly significant effect in favour of variants of CBT for those receiving between one and six sessions (studies 30 and 52*) and, although non-significant, suggested a difference in the same direction in trials where seven to 12 sessions were offered (studies 22, 29, 33, 36 and 45*). However, in trials offering 13–20 sessions (studies 5 and 31*), no difference was observed between groups.†

Trial quality also appeared to affect outcome, with lower-quality trials (scoring 0–19; studies 29, 30, 36 and 52*) demonstrating a highly significant difference in favour of variants of CBT, but higher-quality trials (scoring 20–30; studies 5, 22, 31, 33 and 45*) failing to demonstrate any significant

* See appendix 6 for the references for each study.

† Plots derived from the heterogeneity and sensitivity analyses can be obtained from the author on request.

difference between groups. Again, although a significant difference was demonstrated overall, differences between groups according to the source of recruitment were also suggested, with trials that had recruited volunteers or responders to advertisements (studies 5, 31, 33, 36 and 52*) demonstrating a larger effect in favour of variants of CBT than those trials that had recruited via outpatient clinics and referrals (studies 22, 29, 30 and 45*). In both cases, the differences were non-significant and statistical heterogeneity remained.†

Mean change

Three trials reported the mean change in symptom levels from baseline for this comparison on a total of 130 patients (studies 5, 36 and 45*; see *Figure 42*). None of the effect sizes from individual trials demonstrated any differences between the two types of therapy. The approximate χ^2 test for heterogeneity was non-significant ($\chi^2 = 1.50$, df = 2, $p = 0.47$) and a fixed-effects model was used to pool the data from these trials. The WMD for mean change was 1.37 (95% CI, –1.61 to 4.36) and was non-significant ($z = 0.90$, $p = 0.4$), suggesting no difference in the reduction of symptoms from baseline between the two types of therapy. The small number of trials providing data for this outcome prevented any further exploration of heterogeneity.

Follow-up outcomes

Non-symptomatic at follow-up

A total of three trials reported data for inclusion in the pooled analysis for the number of patients that were non-symptomatic at follow-up (studies 5, 28 and 33*; see *Figure 43*). Data were pooled using a fixed-effects model. All three trials reported data for a follow-up of up to 3 months ($n = 147$) and demonstrated no differences between the two types of therapy, either individually or once pooled (OR = 1.33, 95% CI, 0.66 to 2.67; $z = 0.79$, $p = 0.4$). One small trial (study 33*) reported follow-up for this outcome at 6 months ($n = 23$) and demonstrated a significant difference in favour of variants of CBT at follow-up (OR = 12.15, 95% CI, 2.45 to 60.30; $z = 3.06$, $p = 0.002$). The small number of trials providing data for this outcome prevented any further exploration of heterogeneity.

Mean differences at follow-up

Three trials contributed data on the mean differences between groups at follow-up (studies 22, 29 and 33*). Two trials reported mean differences at a follow-up of up to 2 months ($n = 46$). One trial suggested a significant effect in favour of variants of CBT (study 29*) and the other suggested no difference between the groups (study 22*). Using a fixed-effects model, the pooled estimate demonstrated a significant difference in favour of variants of CBT (WMD = –3.40, 95% CI, –5.56 to –1.25; $z = 3.09$, $p = 0.002$). One small trial reported mean differences data at 6-month follow-up ($n = 17$; study 33*) and suggested a significant difference in favour of variants of CBT (see *Figure 44*). The small number of trials providing data for this outcome prevented any further exploration of heterogeneity.

Mean change at follow-up

Only one trial reported mean change data at 3-month follow-up ($n = 55$; study 5*) and suggested no difference between the two types of therapy (WMD = 0.73, 95% CI, –4.06 to 5.52; $z = 0.30$, $p = 0.8$).

Economic outcomes

There was no economic evidence for this comparison.

Comparison 9: IPT versus ST

Post-treatment outcomes

Recovery

One trial provided data for post-treatment recovery on a total of 48 patients (study 12*). The individual OR for post-treatment recovery demonstrated no difference between IPT and ST (OR = 0.90, 95% CI, 0.25 to 3.27) and was non-significant ($z = –0.17$, $p = 0.9$).

Dropouts

The same trial reported dropouts at post-treatment. The individual OR demonstrated no difference in treatment discontinuation between IPT and ST (OR = 2.05, 95% CI, 0.65 to 6.46) and this result was non-significant ($z = 1.23$, $p = 0.2$).

Mean differences and mean change

No trials reported the mean differences between groups or data on mean change in symptom levels from baseline for this comparison.

Follow-up outcomes

No follow-up data at all was available for this comparison.

* See appendix 6 for the references for each study.

† Plots derived from the heterogeneity and sensitivity analyses can be obtained from the author on request.

Economic outcomes

There was no economic evidence available for this comparison.

Comparison 10: IPT versus treatment as usual/a waiting-list control

Post-treatment outcomes

Recovery

Only one trial provided data for post-treatment recovery on a total of 185 patients (study 44*). The individual OR for post-treatment recovery demonstrated a difference in favour of IPT over treatment as usual (OR = 3.52, 95% CI, 1.91 to 6.51) and was highly significant ($z = 4.02$, $p = 0.00006$).

Dropouts

No trials reported dropouts at post-treatment for this comparison.

Mean differences

The same trial reported mean difference between groups on a total of 185 patients, and the mean difference was significant in favour of IPT. The WMD was –3.21 (95% CI, –5.18 to –1.24) and was statistically significant ($z = 3.20$, $p = 0.001$) suggesting that the reduction in symptoms was greater in those receiving IPT than in those receiving treatment as usual.

Mean change

The mean change in symptom levels from baseline reported, again, by the same trial on a total of 185 patients suggested a possible difference in favour of IPT. The WMD for mean change was 2.09 (95% CI, 0.04 to 4.14) and had borderline significance ($z = 1.99$, $p = 0.05$).

Follow-up outcomes

Non-symptomatic at follow-up

The same trial reported this outcome at 8 months on 185 patients and, again, demonstrated a significant difference in favour of IPT compared with treatment as usual (OR = 3.72, 95% CI, 2.03 to 6.81; $z = 4.26$, $p = 0.00002$).

Mean differences at follow-up

The same trial reported mean difference between groups at 8-month follow-up on a total of 185 patients, which was significant in favour of IPT. The WMD was –3.77 (95% CI, –5.66 to –1.88) and was statistically significant ($z = 3.91$, $p = 0.00009$) suggesting that the greater reduction in symptoms in those receiving IPT was sustained at 8-month follow-up.

Mean change at follow-up

The mean change in symptom levels from baseline to 8-month follow-up reported by the same trial on a total of 185 patients suggested a possible difference in favour of IPT. The WMD for mean change was 2.64 (95% CI, 0.56 to 4.72) and was statistically significant ($z = 2.49$, $p = 0.01$) suggesting that the greater change in symptom levels in those receiving IPT was sustained at 8 months.

Economic outcomes

The associated economic evaluation suggested that IPT was more cost-effective than usual care only if delivered by non-psychiatrists.[60]

Comparison 11: PDT versus ST

Post-treatment outcomes

Recovery

Only one trial provided data for post-treatment recovery on a total of 99 patients (study 28*). The individual OR for post-treatment recovery demonstrated no difference between ST and PDT (OR = 0.83, 95% CI, 0.33 to 2.09) and was non-significant ($z = 0.40$, $p = 0.7$).

Dropouts

The same trial reported dropouts at post-treatment. The individual OR demonstrated no difference in treatment discontinuation between ST and PDT (OR = 1.36, 95% CI, 0.41 to 4.52) and was non-significant ($z = 0.50$, $p = 0.6$).

Mean differences and mean change

No trials reported mean differences between groups or data on mean change in symptom levels from baseline for this comparison.

Follow-up outcomes

Non-symptomatic at follow-up

Only one trial provided data for analysis at 3-month follow-up on a total of 99 patients (study 28*). The individual OR demonstrated no difference between PDT and ST (OR = 1.33, 95% CI, 0.51 to 3.46) and was non-significant ($z = 0.58$, $p = 0.6$).

Mean differences and mean change at follow-up

No trials reported mean differences between groups or data on mean change in symptom levels from baseline to follow-up for this comparison.

Economic outcomes

There was no economic evidence available for this comparison.

* See appendix 6 for the references for each study.

Comparison 12: ST versus treatment as usual/a waiting-list control

Post-treatment outcomes

Recovery

Four trials provided sufficient data for inclusion in the pooled analysis for post-treatment recovery on a total of 118 patients (studies 16, 33, 45 and 50*). None of the individual trials demonstrated differences in recovery between variants of ST and treatment as usual. The approximate χ^2 test for heterogeneity was non-significant ($\chi^2 = 0.77$, df = 3, $p = 0.86$) and a fixed-effects model was used. The pooled OR for recovery was 2.71 (95% CI, 1.19 to 6.16) suggesting that the odds of recovery for those receiving ST were over two and half times that of those receiving treatment as usual. The result was statistically significant ($z = 2.38$, $p = 0.02$; see *Figure 45*). The pooled RR for recovery using a random-effects model was 1.58 (95% CI, 1.03 to 2.41). The small number of trials providing data for this outcome prevented any further exploration of heterogeneity.

Dropouts

Four trials reported dropouts at post-treatment on a total of 146 patients (studies 16, 30, 33 and 45*). None of the individual trials demonstrated a difference between ST and treatment as usual. The approximate χ^2 test for heterogeneity was non-significant ($\chi^2 = 4.82$, df = 3, $p = 0.19$) and a fixed-effects model resulted in a pooled OR of 1.74 (95% CI, 0.68 to 4.46). This was non-significant ($z = 1.16$, $p = 0.2$) suggesting no differences in treatment discontinuation between ST and treatment as usual (see *Figure 46*). The small number of trials providing data for this outcome prevented any further exploration of heterogeneity.

Mean differences

Four trials reported mean differences between groups for this comparison on a total of 123 patients (studies 30, 33, 45 and 57*). One of the mean differences from individual trials was in favour of ST (study 45*) whilst the remainder demonstrated no differences between groups (studies 30, 33 and 57*). The approximate χ^2 test for heterogeneity was non-significant ($\chi^2 = 1.18$, df = 3, $p = 0.76$) and, therefore, a fixed-effects model was used to pool the data from these trials. The SMD was –0.42 (95% CI, –0.78 to –0.06) in favour of ST and the result was significant ($z = 2.29$, $p = 0.02$), suggesting that the reduction in symptoms was greater in those receiving ST than in those receiving treatment as usual (see *Figure 47*). The small number of trials providing data for this outcome prevented any further exploration of heterogeneity.

Mean change

Two trials reported mean change in symptom levels from baseline on a total of 79 patients (studies 45 and 57*). Neither of the effect sizes from individual trials demonstrated any differences between ST and treatment as usual. The approximate χ^2 test for heterogeneity was non-significant ($\chi^2 = 0.00$, df = 1, $p = 0.96$) and thus a fixed-effects model was used to pool the data. The SMD for mean change was –0.01 (95% CI, –0.45 to 0.43) and was non-significant ($z = 0.05$, $p = 1$), demonstrating no difference in the reduction of symptoms from baseline between ST and treatment as usual (see *Figure 48*). The small number of trials providing data for this outcome prevented any further exploration of heterogeneity.

Follow-up outcomes

Non-symptomatic at follow-up

No data for this outcome at follow-up were available.

Mean differences at follow-up

One trial reported mean differences data at 2-month follow-up ($n = 22$; study 57*) and suggested no difference between ST and treatment as usual (SMD = –0.27, 95% CI, –1.11 to 0.57; $z = 0.63$, $p = 0.5$). A second trial reported mean differences data at a follow-up of 6 months ($n = 17$; study 33*) and also suggested that there was no difference between ST and treatment as usual (SMD = –2.01, 95% CI, –3.27 to –0.75; $z = 3.13$, $p = 0.002$).

Mean change at follow-up

Only one trial reported mean change data at 2-month follow-up ($n = 22$; study 57*) and suggested that there was no difference between the two types of therapy (WMD = –4.30, 95% CI, –17.05 to 8.45; $z = 0.66$, $p = 0.5$).

Economic outcomes

No cost-effectiveness analyses were available for any of the included trials in this comparison. One other cost-effectiveness analysis linked to an excluded trial by Katon (1995; see appendix 7 for the reference for this study), and referred to earlier in comparison 5, concluded that ST offered within a collaborative care model was more costly

* See appendix 6 for the references for each study.

but achieved greater success than treatment as usual. This translated into a modest cost-effectiveness advantage.

Other sources of bias

Funnel plots were conducted for two dichotomous outcomes for which there were adequate data. These are scatter plots of treatment effect sizes estimated from individual studies on the horizontal axis against a measure of study size on the vertical axis. Each circle represents a study and the dotted vertical line indicates the overall (pooled) effect size for the outcome.

The graphs show a plot of standard error (SE) versus effect size. The statistical power of a study is determined by the total sample size and the number of events, thus it is sensible to base measures on the SE or variance of the effect estimate, rather than the total sample size. Using the SE emphasises the differences between studies of smaller size, for which biases are most likely to operate. A funnel plot is based on the fact that the precision in the estimation of the underlying treatment effect will increase as the sample size of component studies increases. Effect estimates from small studies will, therefore, scatter more widely at the bottom of the graph, with the spread narrowing among larger studies.

The first funnel plot was of the data from comparison 1 (all variants of psychotherapy versus treatment as usual) for the outcome of recovery. As can be seen in *Figure 49*, small- and medium-sized trials comparing any variant of psychotherapy versus treatment as usual/a waiting-list control, where no difference was observed or where active treatment had a worse outcome, appear to be missing, suggesting the likelihood of bias towards positive trials reporting results in favour of the intervention group. A similar effect can be seen in the other four funnel plots (*Figures 50–53*).

TABLE 1 *Populations studied in the included trials*

	Number of trials	%	Study number*
Country			
USA	50	79	2–27, 29, 31–39, 41–44, 48, 51–53, 56, 57, 59, 61–63
UK	8	13	1, 40, 45–47, 49, 54, 55
Canada	3	5	28, 30, 50
Australia	2	3	58, 60
Setting			
University psychology departments	43	68	2, 3, 5–7, 9, 14–16, 18–24, 26, 27, 29, 32–39, 43, 48–53, 55–63
Outpatient clinics	13	21	1, 4, 8, 10–13, 17, 28, 30, 31, 41, 42
Primary care	7	11	25, 40, 44–47, 54
Recruitment			
Outpatient referrals/attenders	20	32	1, 8, 10, 12, 13, 22, 25, 29, 30, 40–42, 44–47, 49, 50, 54, 55
Media advertisements	22	35	6, 11, 15, 16, 18, 20, 26–28, 32–34, 37–39, 43, 48, 58–60, 62, 63
Student volunteers	11	17	3, 7, 9, 14, 19, 24, 35, 51–53, 56
Combination of methods	10	16	2, 4, 5, 17, 21, 23, 31, 36, 57, 61
Diagnosis for inclusion			
MDD	24	38	1, 4, 5, 11–13, 20, 23, 25, 26, 31, 32, 34, 38–40, 44–47, 49, 54, 58, 61
Mild/moderate depression	11	17	3, 7, 9, 10, 14, 21, 35, 36, 53, 57, 59
Depression of unspecified severity	24	38	2, 8, 15, 16, 18, 19, 22, 24, 27–30, 33, 37, 41–43, 50–52, 56, 60, 62, 63
Any level of depression	4	6	6, 17, 48, 55
Diagnostic procedure			
Diagnostic research instrument + depression severity measure	31	49	1, 4, 5, 11–13, 17, 20, 23, 26, 28, 31–34, 36, 38–42, 44, 46–49, 54, 58, 59, 61, 63
Diagnostic research instrument	4	6	6, 10, 45, 55
Clinical interview + depression severity measure	7	11	2, 15, 22, 29, 37, 43, 50
Depression severity measure	21	33	3, 7–9, 14, 16, 18, 19, 21, 24, 25, 27, 30, 35, 51–53, 56, 57, 60, 62
Exclusion criteria†			
On medication for depression	24	37	4, 5, 10–12, 17, 20, 23, 26, 28, 31–36, 38, 39, 53, 56, 58, 59, 62, 63
Currently in psychotherapy	24	37	1, 4, 5, 12, 16, 17, 22, 23, 26, 27, 32–35, 37, 38, 44, 49, 53, 54, 56, 58, 59, 61
Suicidal ideation/attempts	23	37	5, 8–10, 13, 16, 17, 21–23, 25–27, 29, 35, 37, 42, 43, 45, 50, 57, 59, 60
No exclusion criteria reported	10	16	2, 7, 9, 18, 19, 21, 24, 40, 51, 52

* *See appendix 6 for the references for each study*

† *Some trials used more than one exclusion criterion, and the total percentage, therefore, exceeds 100%*

TABLE 2 *Treatments studied in the included trials. All trials are included in a minimum of two categories*

Category	Models included	Number of trials	%	Study number*
Cognitive behavioural	CBT (to include CT and BT)	53	84	1, 2, 4–6, 8, 9, 11, 13, 15–18, 21–23, 25–43, 45–56, 58–63
	CBT techniques	6	10	3, 7, 9, 14, 19, 24
Interpersonal	IPT	8	13	1, 11–13, 22, 32, 44, 49
Psychodynamic	PDT	10	16	1, 4, 10, 11, 17, 26–28, 42, 49
Supportive	Client-centred therapy	6	10	20, 29, 30, 36, 50, 56
	Gestalt therapy	3	5	5, 20, 57
	Non-specific therapy	12	19	10, 12, 13, 15, 16, 28, 31, 33, 45, 48, 52, 57
Treatment as usual	Waiting list	21	33	2, 6–9, 16, 21, 22, 24, 30, 33, 34, 38, 40, 43, 48, 50, 51, 53, 58, 60
	Treatment as usual	5	8	25, 44–46, 54
	No treatment	4	6	3, 14, 19, 35
Therapy format	Individual therapy	29	46	1, 3, 4, 9, 12–14, 17, 19, 20, 23–25, 28, 29, 31, 35, 44–46, 48, 49, 51–54, 56, 60, 61
	Group therapy	26	41	2, 5, 7, 8, 10, 11, 15, 16, 18, 21, 22, 26, 27, 30, 32–34, 36–39, 42, 50, 57, 58, 62
	Both	8	13	6, 40, 41, 43, 47, 55, 59, 63
Primary objective	Evaluation of one model or technique	17	29	2, 14, 16, 19, 21, 24, 25, 30, 33, 35, 42, 44, 46, 48, 51, 54, 58
	Comparison of two or more models or techniques	24	38	1, 3, 4, 7–13, 15, 17, 18, 22, 27, 28, 31, 36, 45, 49, 50, 52, 56, 60
	Comparison of therapy format (group versus individual)	8	13	6, 40, 41, 43, 47, 55, 59, 63
	Model dismantling analysis (efficacy of different components of therapy)	11	17	23, 26, 29, 34, 37–39, 53, 57, 61, 62
	Analysis of therapeutic process	1	2	20
	Identification of predictive factors for positive outcome	2	3	5, 32

* *See appendix 6 for the references for each study*

TABLE 3 *Use of antidepressant intervention in the included trials*

	Number of trials	%	Study number*
Pharmacotherapy arm(s)			
Pharmacotherapy versus psychotherapy	5	8	13, 28, 31, 44, 45
Combination therapy	5	8	4, 10–12, 41
Psychotherapy trials			
Subjects on antidepressants excluded	18	29	5, 17, 20, 23, 26, 32–36, 38, 39, 53, 56, 58, 59, 62, 63
Some subjects in whole sample on antidepressants	10	16	1, 25, 27, 30, 42, 46, 47, 49, 54, 60
Some subjects in control group on antidepressants	1	2	42
Not reported	24	38	2, 3, 6–9, 14–16, 18, 19, 21, 22, 24, 29, 37, 43, 48, 50–52, 55, 57, 61
Pharmacotherapy/psychotherapy trials			
Psychotherapy alone	5	8	11–13†, 28, 31
Antidepressant placebo given with psychotherapy	3	5	4, 10, 13†
Some subjects in psychotherapy or control groups on antidepressants	3	5	41, 44, 45

* *See appendix 6 for the references for each study*
† *This trial appears twice as it included three arms of psychotherapy, one of which was a combination of ST and placebo antidepressant*

TABLE 4 *Application of psychotherapy in the included trials*

	Number of trials	%	Study number*
Therapists' qualifications			
Qualified specialist (psychiatrist, psychologist)	16	25	1, 10–14, 16, 23, 25, 28, 31, 44, 47, 49, 54, 63
Advanced level clinical/counselling psychology students	23	37	4, 6, 9, 17, 26, 27, 29, 30, 32, 33, 36–40, 42, 48, 51, 52, 55–58
Clinical/counselling psychology students	11	17	2, 3, 15, 21, 22, 34, 35, 50, 53, 59, 61
Range of qualifications	7	11	5, 20, 41, 43, 45, 46, 62
Not reported	6	10	7, 8, 18, 19, 24, 60
Specialist pre-trial training			
Training of ≥ 3 months	6	10	1, 13, 20, 23, 26, 55
Training of < 3 months	8	13	3, 9, 14, 15, 29, 31, 38, 44
Training of unspecified length/content	15	24	6, 16, 21, 22, 25, 28, 33, 34, 39, 43, 48, 49, 58, 59, 61
No specialist training	34	54	2, 4, 5, 7, 8, 10–12, 17–19, 24, 27, 30, 32, 35–37, 40–42, 45–47, 50–54, 56, 57, 60, 62, 63
Monitoring of therapy			
Tapes assessed by independent raters	24	38	3, 5, 8, 9, 13, 15, 17, 20, 22, 23, 25, 26, 28, 29, 31, 36, 39, 46, 49, 50, 54, 57, 58, 62
Tapes discussed in supervision	13	21	1, 6, 16, 32–34, 37, 38, 41, 43, 55, 59, 61
One-way mirror observation	2	3	35, 44
No monitoring	24	38	2, 4, 7, 10–12, 14, 18, 19, 21, 24, 27, 35, 40, 42, 45, 47, 48, 51–53, 56, 60, 63
Number of sessions			
One–six	22	35	3, 7–9, 14, 16, 18, 19, 24, 25, 27, 30, 39, 46, 48, 51–53, 56–59
Seven–12	26	41	2, 4, 6, 15, 21, 22, 26, 28, 29, 32–38, 42, 43, 45, 47, 50, 55, 60–63
13–20	13	21	5, 10–13, 17, 20, 23, 31, 40, 41, 44, 54
Contrasting numbers of sessions	2	3	1, 49

* *See appendix 6 for the references for each study*

TABLE 5 *Outcome measures used in the included trials*

Outcome	Instruments used	Number of trials	%	Study number*
Depression levels	BDI	53	84	1–9, 11, 13–24, 26–35, 37–44, 46–50, 53–55, 58–60, 62, 63
	MMPI	21	33	2, 5, 15, 16, 22, 26, 31, 33, 37–39, 41–43, 51–53, 57, 59, 61, 63
	PES	16	25	2, 15, 16, 21, 26, 29, 36–39, 43, 56, 60–63
	Hopkins SCL/SCL-90-revised	11	17	1, 4, 11–13, 20, 25, 39, 48, 49
	LDACL	7	11	2, 4, 28, 36, 43, 56, 57
	ZSRDS	7	11	18, 24, 30, 36, 43, 51, 58
	RGSD	5	8	4, 12, 26, 38, 39
	HRSD (clinician-rated)	25	40	4, 5, 8, 11–13, 17, 23, 26, 29, 31, 32, 34, 38, 39, 41, 44–46, 48, 50, 54, 60, 62, 63
General symptomatology	Global Assessment Scale	3	5	13, 26, 38
Social functioning	Social Adjustment Scale	7	11	4, 12, 13, 39, 49, 55, 63
Cognitive-behavioural functioning	Dysfunctional Attitude Scale	5	8	13, 15, 39, 62, 63
	Self-esteem Inventory	4	6	1, 20, 22, 30
	Wolpe–Lazarus Assertiveness Scale	2	3	4, 37
	Rathus Assertiveness Schedule	2	3	27, 42
Client satisfaction with therapy	Barrett–Lennard Relationship Inventory	4	6	3, 5,14, 20
	Domains of Satisfaction Scale	2	3	6, 42

* *See appendix 6 for the references for each study*

PES, Pleasant Events Schedule; SCL, Symptom Checklist; LDACL, Lubin Depression Adjective Checklist; ZSRDS, Zung Self-rating Depression Scale; RGSD, Raskin Global Scale for Depression

TABLE 6 *Randomisation, allocation concealment and blinding in the included trials*

	Number of trials	%	Study number*
Study design			
RCT	55	87	1–6, 8–23, 25, 27–35, 37–40, 42–50, 53–55, 57–63
CCT (randomisation not used)	8	13	7, 24, 26, 36, 41, 51, 52, 56
Allocation procedure of RCTs			
Randomisation conducted by independent research staff	2	3	45, 49
Computer-generated	2	3	13, 25
Table of random numbers	5	8	17, 21, 31, 40, 47
Sealed envelopes	1	2	45
Blocked	2	3	5, 29
Procedure not described	45	71	1–4, 6, 8–12, 14–16, 18–20, 22, 23, 27, 28, 30, 32–35, 37–39, 42–44, 46, 48–50, 53–55, 57–63
Allocation procedure of CCTs			
Simple rotation	1	2	56
Coin tossing	1	2	36
Subjects matched	3	5	7, 51, 52
Procedure not described	3	5	24, 26, 41
Stratification of RCTs			
Stratification	5	9	1, 14, 18, 25, 46
Pre-randomisation matching	9	14	8, 16, 17, 20, 21, 23, 49, 50, 58
Blinding			
Subjects blind to therapeutic model	4	6	3, 4, 10, 14
Not described	59	94	1, 2, 5–9, 11–13, 15–63
Blinding at assessment interviews			
Blind raters used	21	33	4, 11–13, 17, 23, 25, 26, 34, 38, 39, 44–46, 48, 50, 54, 55, 61–63
Independent raters used	6	10	1, 6, 28, 31, 32, 41

* *See appendix 6 for the references for each study*

TABLE 7 *Patient follow-up in the included trials*

	Number of trials	%	Study number*
Attrition rate (%) post-treatment†			
No dropouts	4	6	43, 48, 51, 59
1–10%	13	21	14, 21, 23–25, 27, 28, 34–36, 45, 47, 52
11–25%	19	30	2, 11, 15, 17, 20, 26, 29, 31, 33, 38, 39, 41, 44, 54, 57, 60, 62, 63
> 25%	13	21	1, 4, 5, 12, 13, 16, 18, 30, 40, 42, 46, 49, 58
Not reported	14	22	3, 6–10, 19, 22, 32, 37, 50, 53, 55, 56
Final follow-up assessment			
≤ 3 months post-treatment	25	40	3, 8, 9, 14–16, 18, 21, 22, 26, 27, 30, 35, 37, 41, 42, 48, 50–53, 57, 61–63
4–6 months post-treatment	10	16	4, 6, 20, 24, 29, 39, 43, 47, 55, 58
> 6 months post-treatment	20	32	1, 2, 5, 7, 11–13, 17, 23, 25, 28, 33, 34, 38, 40, 44, 46, 49, 54, 60
Follow-up not conducted	8	13	19, 31, 32, 36, 45, 47, 56, 59
Maintenance treatment given	2	3	4, 10

* *See appendix 6 for the references for each study*
† *The mean attrition rate in the 49 trials reporting dropouts before and during treatment was 17% (SD 13.65)*

TABLE 8 *Reporting and analysis in the included trials*

	Number of trials	%	Study number*
Method of analysis			
ITT analysis/complete data sets	11	17	23, 25, 36, 40, 43–45, 47, 48, 51, 59
Endpoint analysis	7	11	4, 5, 10–13, 31
Completers only	32	51	1, 2, 6–9, 14–17, 20, 21, 26, 27, 29, 30, 33–35, 37, 39, 41, 42, 46, 49, 54, 55, 57, 58, 61–63
Dropouts replaced	3	5	28, 56, 60
Insufficient data to identify method	10	16	3, 18, 19, 22, 24, 32, 38, 50, 52, 53
Identification of main outcomes			
Outcomes specified *a priori*	20	32	4–6, 13, 14, 20, 22, 23, 28, 30, 35, 37–39, 49, 50, 54–56, 63
Outcomes stated	38	60	1–3, 7–9, 11, 12, 15–19, 21, 24–26, 29, 31–34, 36, 40–48, 57–62
Outcomes unclear	5	8	10, 27, 51–53
Reporting of data†			
Measures of variability (SDs)	46	73	1, 3–7, 9, 13, 14, 16, 18, 20, 22–26, 29–31, 33–41, 43–49, 51–53, 55–57, 59, 60, 62, 63
Graphs	25	40	1, 4, 6, 10–13, 16, 18, 25, 27, 28, 30, 31, 34, 37, 41, 43, 44, 47–49, 52, 56

* *See appendix 6 for the references for each study*
† *The total percentage does not total 100% because some trials reported more than one aspect of statistical information*

TABLE 9 *Treatment outcomes used in the included trials*

Outcome	Time of assessment	Number of trials	%	Study number*†
Recovery	Post-treatment	32	51	4, 5, 11–13, 16, 17, 23, 25, 26, 28, 29, 31, 33–35, 40, 41, 44–50, 52, 54–56, 58, 59, 63 (*5, 17, 25, 26, 40, 44–46, 49, 58*)
Non-symptomatic	Final follow-up	14	22	5, 6, 17, 23, 25, 28, 33, 35, 44, 48, 49, 54, 55, 63 (*5, 17, 25, 44, 49*)
Reported dropouts	Post-treatment	38	60	4–7, 11–14, 16, 17, 21, 23, 26, 28–31, 33–36, 38, 40–43, 45–50, 52, 54, 56, 58, 60, 63 (*5, 17, 26, 40, 45, 46, 49, 58*)
Mean change from baseline	Post-treatment	10	16	5, 17, 21, 36, 44–46, 49, 57, 58 (*5, 17, 44–46, 49, 57, 58*)
	Final follow-up	6	10	5, 17, 25, 44, 46, 57 (*5, 17, 25, 44, 46, 57*)
Mean differences between groups	Post-treatment	40	63	3–7, 9, 13, 14, 17, 22, 23, 25, 26, 29–31, 33–36, 38–41, 43–48, 51–53, 55, 57–60, 62, 63 (*5, 17, 25, 26, 40, 44–46, 57, 58*)
	Final follow-up	26	41	3, 5–7, 9, 14, 17, 22, 23, 25, 26, 29, 33, 39, 40, 43, 44, 46–48, 53, 55, 57, 60, 62, 63 (*5, 17, 25, 26, 40, 44, 46, 57*)
Dichotomous and continuous symptom outcomes reported	Post-treatment	26	41	4–6, 13, 17, 23, 25, 26, 29, 31, 33–35, 40, 41, 44–49, 52, 55, 58, 59, 63 (*5, 17, 25, 26, 40, 44–46, 58*)
	Final follow-up	10	16	5, 6, 17, 23, 25, 33, 44, 48, 55, 63 (*5, 17, 25, 44*)
Excluded from meta-analysis due to inappropriate comparison or insufficient data		13	19	1, 2, 8, 10, 15, 18–20, 24, 27, 32, 37, 61

* *See appendix 6 for the references for each study*
† *The italicised study numbers in parentheses indicate those for which additional data were submitted by authors*

TABLE 10 *Grouping of QRS items*

Item number	Items	Group
1	Objectives and specification of main outcomes *a priori*	3
2	Sample size (number per group)	3
3	Planned duration of trial including follow-up	4
4	Power calculation	3
5	Method of allocation	3
6	Concealment of allocation	3
7	Clear description of treatment and adjunctive treatment	2
8	Blinding of subjects	3
9	Source of subjects described and representative sample recruitment	1 + 4
10	Use of diagnostic criteria or clear specification of inclusion criteria	1 + 4
11	Record of exclusion criteria and number of exclusions/refusals	1 + 4
12	Description of sample demographics	1 + 4
13	Blinding of assessor	3
14	Assessment of compliance with experimental treatments	2
15	Details on side-effects	2
16	Record of number and reasons for withdrawals by group	3 + 4
17	Outcome measures described or use of validated instruments	3
18	Information on comparability/adjustment for differences in analysis	3
19	Inclusion of withdrawals in analysis	3 + 4
20	Presentation of results with inclusion of data for re-analysis	3
21	Appropriate statistical analysis	3
22	Conclusions justified	3 + 4
23	Declaration of interests (e.g. source of funding)	3

TABLE 11 *Quality rating of trials according to groups. A list of scores for all individual QRS items relating to each trial is attached in appendix 10*

Group	Items from QRS	Score	Number of trials	%	Study number*
1 + 4 Patient group and external validity	9, 10, 11, 12	0–2	19	30	2, 3, 7–9, 14–16, 19, 21, 24, 30, 47, 51–53, 55–57
		3–5	26	41	5, 6, 18, 22, 27–29, 32–37, 39–43, 48–50, 59–63
		6–8	18	29	1, 4, 10–13, 17, 20, 23, 25, 26, 31, 38, 44–46, 54, 58
2 Intervention	7, 14, 15	0–2	43	68	1, 2, 7–11, 13–19, 22–24, 26, 29, 30, 32–34, 37–39, 42, 44, 46–53, 55–61
		3–4	19	30	3, 5, 6, 12, 20, 21, 25, 27, 28, 31, 35, 36, 40, 41, 43, 45, 54, 62, 63
		5–6	1	2	4
3 Internal validity	1, 2, 4, 5, 6, 8, 13, 17, 18, 20, 21, 23	0–8	26	41	1–3, 7, 9, 15, 18, 19, 21, 24, 27, 32, 36, 37, 41, 42, 47, 51–53, 56–61
		9–15	35	56	5, 6, 8, 10–12, 14, 16, 17, 20, 22, 23, 25, 26, 28–31, 33–35, 38–40, 43–46, 48–50, 54, 55, 62, 63
		16–24	2	3	4, 13
3 + 4 Internal validity and external validity	16, 19, 22	0–2	38	60	1–3, 6–9, 11, 15, 16, 18, 19, 22, 24, 27–30, 32, 34–38, 42, 46, 49, 50–57, 59, 62, 63
		3–4	22	35	4, 5, 10, 12–14, 17, 20, 21, 23, 26, 33, 39–41, 43–45, 47, 58, 60, 61
		5–6	3	5	25, 31, 48
4 External validity	3	0	23	37	3, 7–9, 14, 15, 18, 19, 30, 32, 35–38, 41, 42, 50–53, 56, 57, 59
		1	16	25	16, 21, 22, 24, 26, 27, 29, 31, 43, 45, 48, 54, 58, 61-63
		2	24	38	1, 2, 4–6, 10–13, 17, 20, 23, 25, 28, 33, 34, 39, 40, 44, 46, 47, 49, 55, 60
1, 2, 3 + 4 Overall score of trials	All items	0–9	8	13	7, 15, 19, 24, 51, 52, 56, 57
		10–19	24	38	1–3, 8, 9, 14, 16, 18, 21, 27, 29, 30, 32, 35–37, 41, 42, 47, 50, 53, 58, 59, 61
		20–29	28	44	5, 6, 10–12, 17, 20, 22, 26, 28, 31, 33, 34, 38–40, 43–46, 48, 49, 53–55, 60, 62, 63
		≥ 30	4	6	4, 13, 23, 25

* *See appendix 6 for the references for each study*

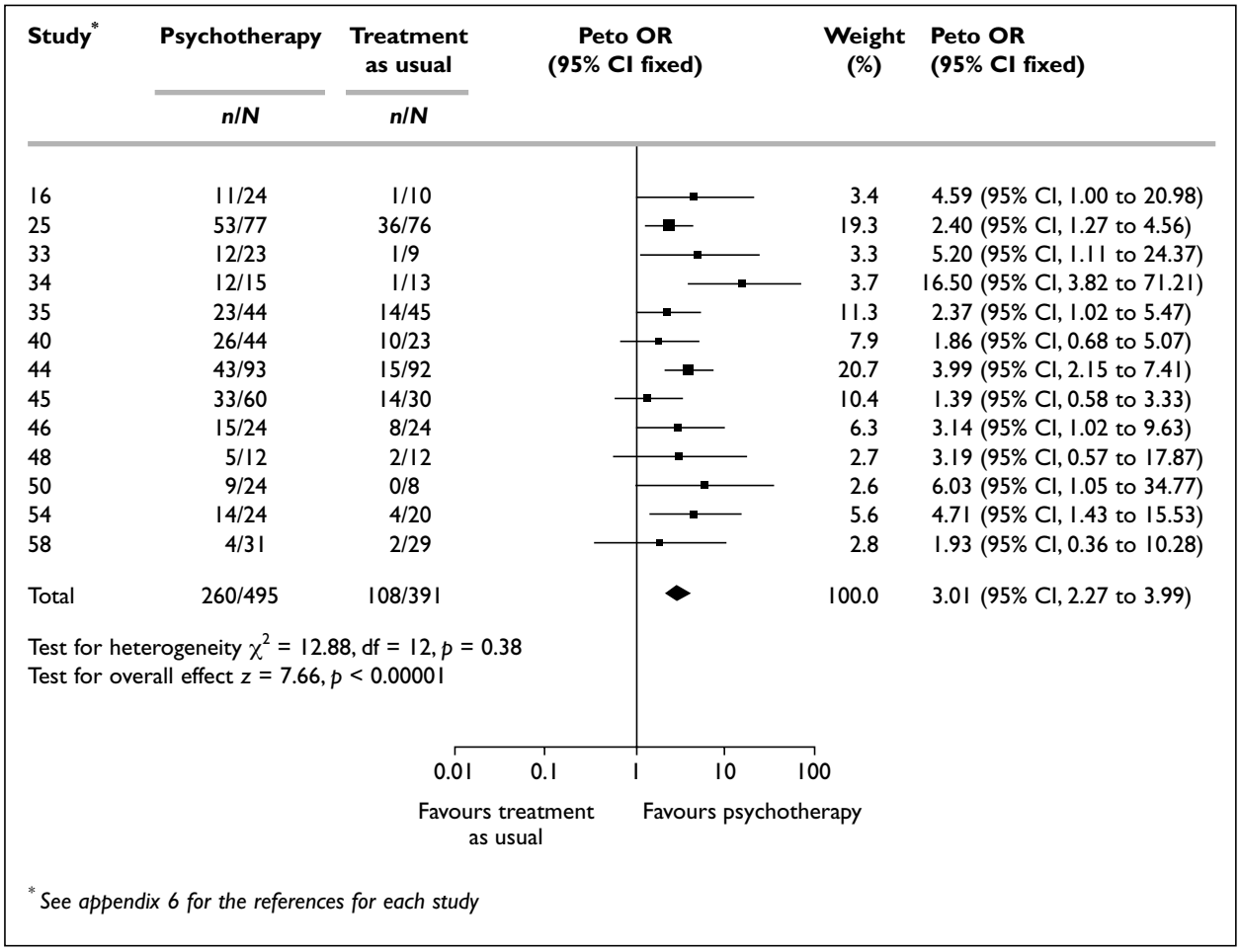
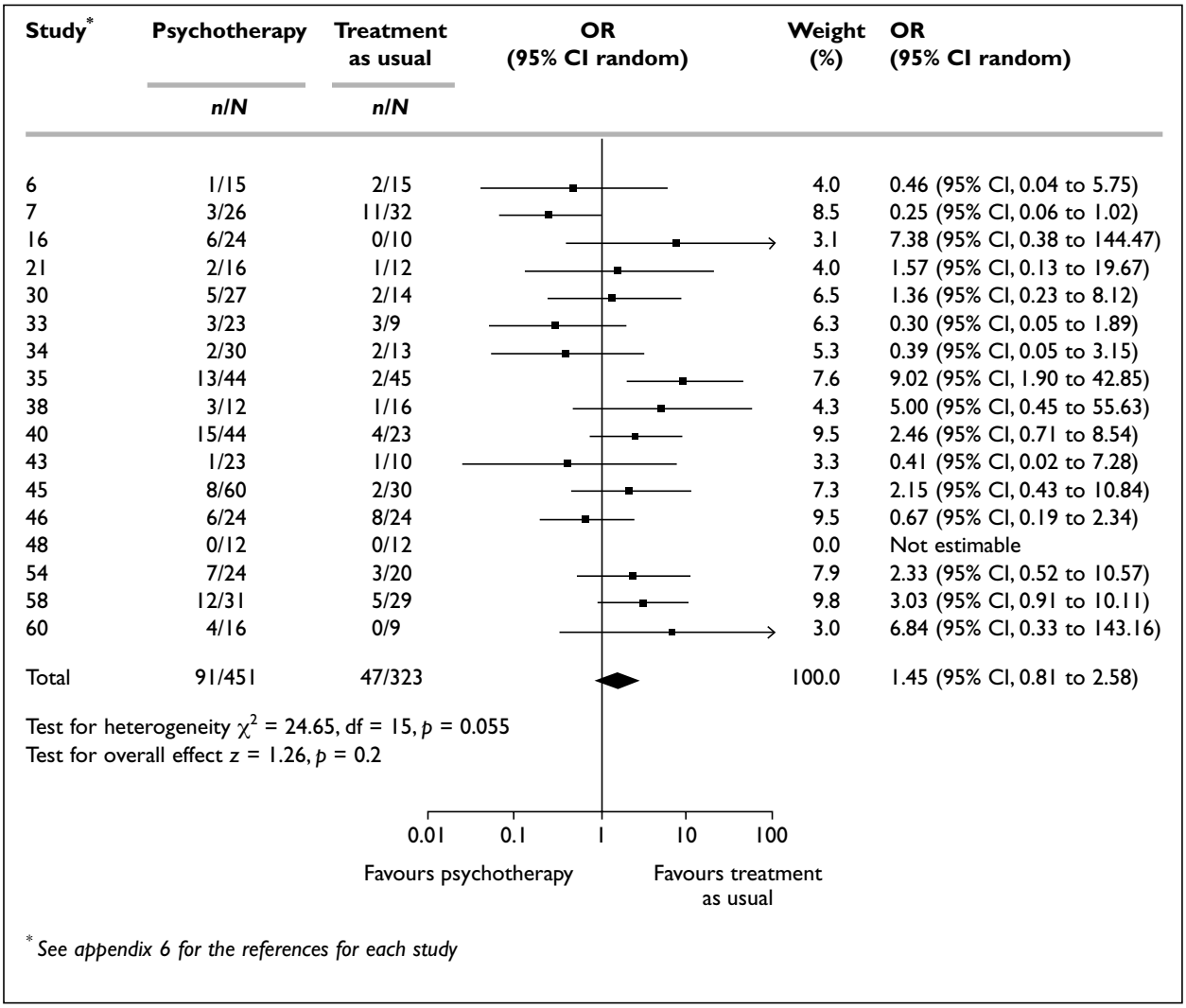


FIGURE 1 *Recovery at post-treatment for all variants of psychotherapy versus treatment as usual/a waiting-list control*

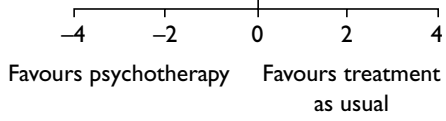


* *See appendix 6 for the references for each study*

FIGURE 2 *Reported dropouts at post-treatment for all variants of psychotherapy versus treatment as usual/a waiting-list control*

Study*	Psychotherapy		Treatment as usual		SMD (95% CI random)	Weight (%)	SMD (95% CI random)
	n	Mean (SD)	*n*	Mean (SD)			
3	10	4.10 (3.54)	10	14.70 (5.08)		3.3	–2.32 (95% CI, –3.51 to –1.13)
6	13	15.08 (10.43)	11	18.18 (11.29)		4.6	–0.28 (95% CI, –1.08 to 0.53)
7	23	8.74 (6.49)	21	17.76 (9.27)		5.2	–1.12 (95% CI, –1.76 to –0.48)
9	19	14.11 (4.51)	19	14.86 (4.89)		5.2	–0.16 (95% CI, –0.79 to 0.48)
14	10	9.60 (8.36)	10	12.00 (7.69)		4.3	–0.29 (95% CI, –1.17 to 0.60)
25	71	26.70 (14.60)	69	25.60 (13.70)		6.3	0.08 (95% CI, –0.25 to 0.41)
30	22	38.96 (7.25)	12	55.00 (9.81)		4.4	–1.91 (95% CI, –2.76 to –1.05)
33	11	9.82 (4.71)	6	21.00 (6.27)		3.2	–2.01 (95% CI, –3.27 to –0.75)
34	14	6.57 (3.29)	11	24.73 (7.76)	←	3.2	–3.09 (95% CI, –4.31 to –1.87)
35	31	6.06 (5.60)	43	11.83 (8.90)		5.8	–0.74 (95% CI, –1.22 to –0.26)
38	9	20.89 (12.12)	15	21.87 (8.20)		4.5	–0.10 (95% CI, –0.92 to 0.73)
40	27	16.81 (10.58)	23	22.78 (10.96)		5.5	–0.55 (95% CI, –1.11 to 0.02)
43	12	7.30 (6.40)	10	21.70 (9.00)		3.8	–1.80 (95% CI, –2.83 to 0.78)
44	93	12.08 (6.68)	92	15.29 (6.97)		6.4	–0.47 (95% CI, –0.76 to –0.18)
45	23	6.74 (6.07)	28	8.75 (7.47)		5.5	–0.29 (95% CI, –0.84 to 0.27)
46	18	17.70 (10.00)	16	22.70 (11.20)		5.0	–0.46 (95% CI, –1.14 to 0.22)
48	12	11.64 (8.20)	12	18.50 (9.32)		4.5	–0.75 (95% CI, –1.59 to 0.08)
51	11	60.90 (9.90)	11	82.80 (9.50)		3.6	–2.17 (95% CI, –3.27 to –1.08)
53	7	5.60 (4.70)	7	20.10 (5.80)	←	2.5	–2.57 (95% CI, –4.11 to –1.04)
57	11	60.80 (15.70)	11	61.30 (16.30)		4.5	–0.03 (95% CI, –0.87 to 0.81)
58	19	16.79 (9.30)	23	22.79 (9.80)		5.3	–0.61 (95% CI, –1.24 to 0.01)
60	8	9.00 (6.82)	9	21.44 (5.52)		3.3	–1.92 (95% CI, –3.12 to –0.71)
Total	474		469		◆	100.0	–0.90 (95% CI, –1.21 to –0.60)

Test for heterogeneity $\chi^2 = 89.26$, df = 21, $p < 0.00001$
Test for overall effect $z = 5.77$, $p < 0.00001$



* *See appendix 6 for the references for each study*

FIGURE 3 *Mean differences post-treatment for all variants of psychotherapy versus treatment as usual/a waiting-list control*

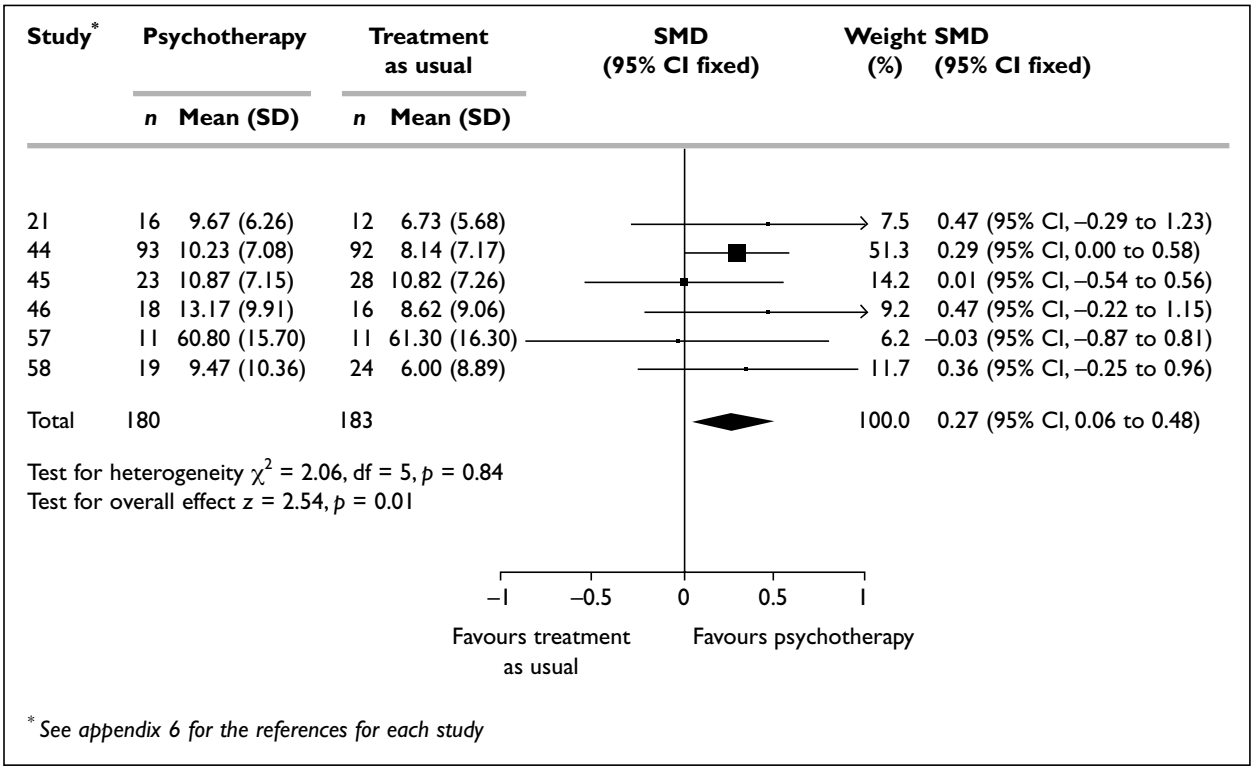


FIGURE 4 *Mean change from baseline at post-treatment for all variants of psychotherapy versus treatment as usual/a waiting-list control*

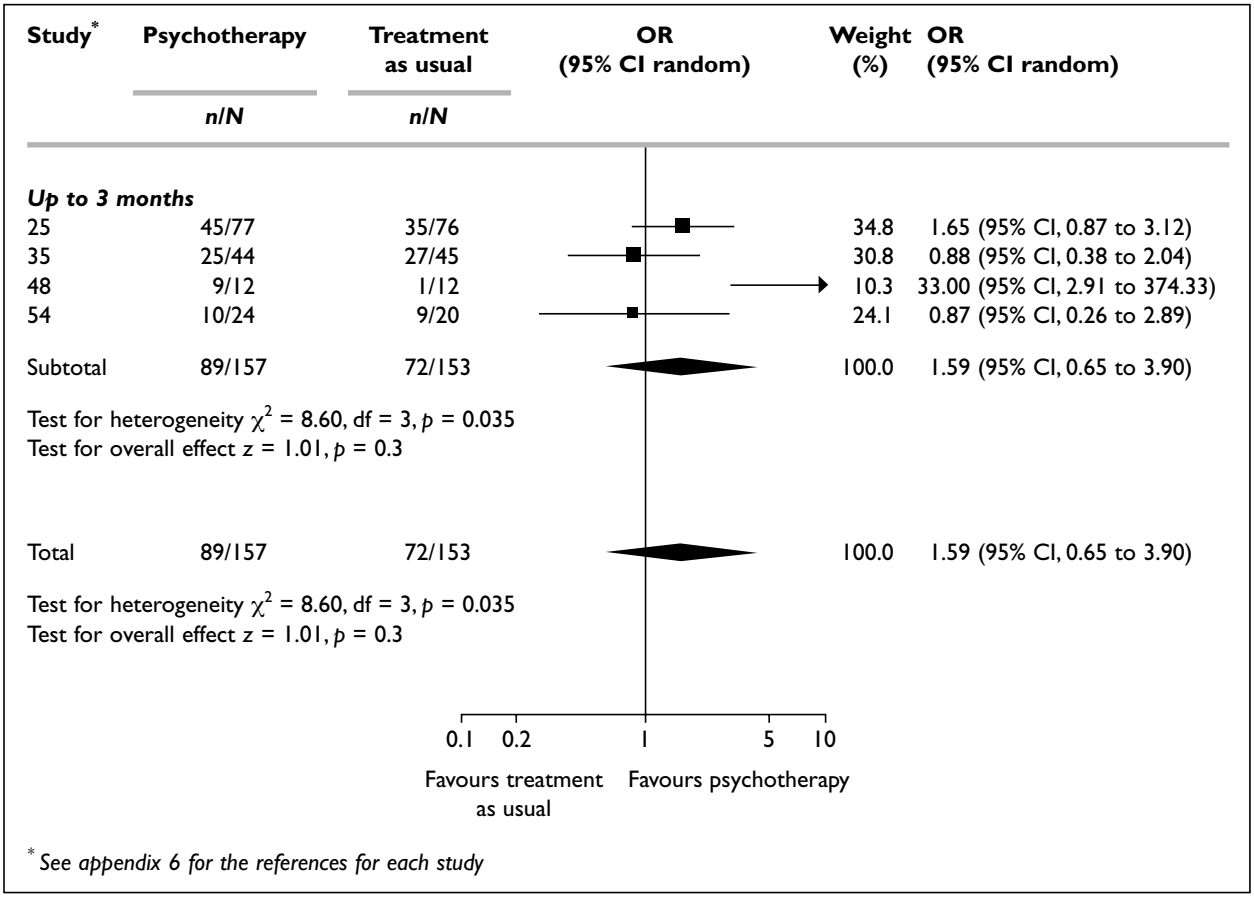
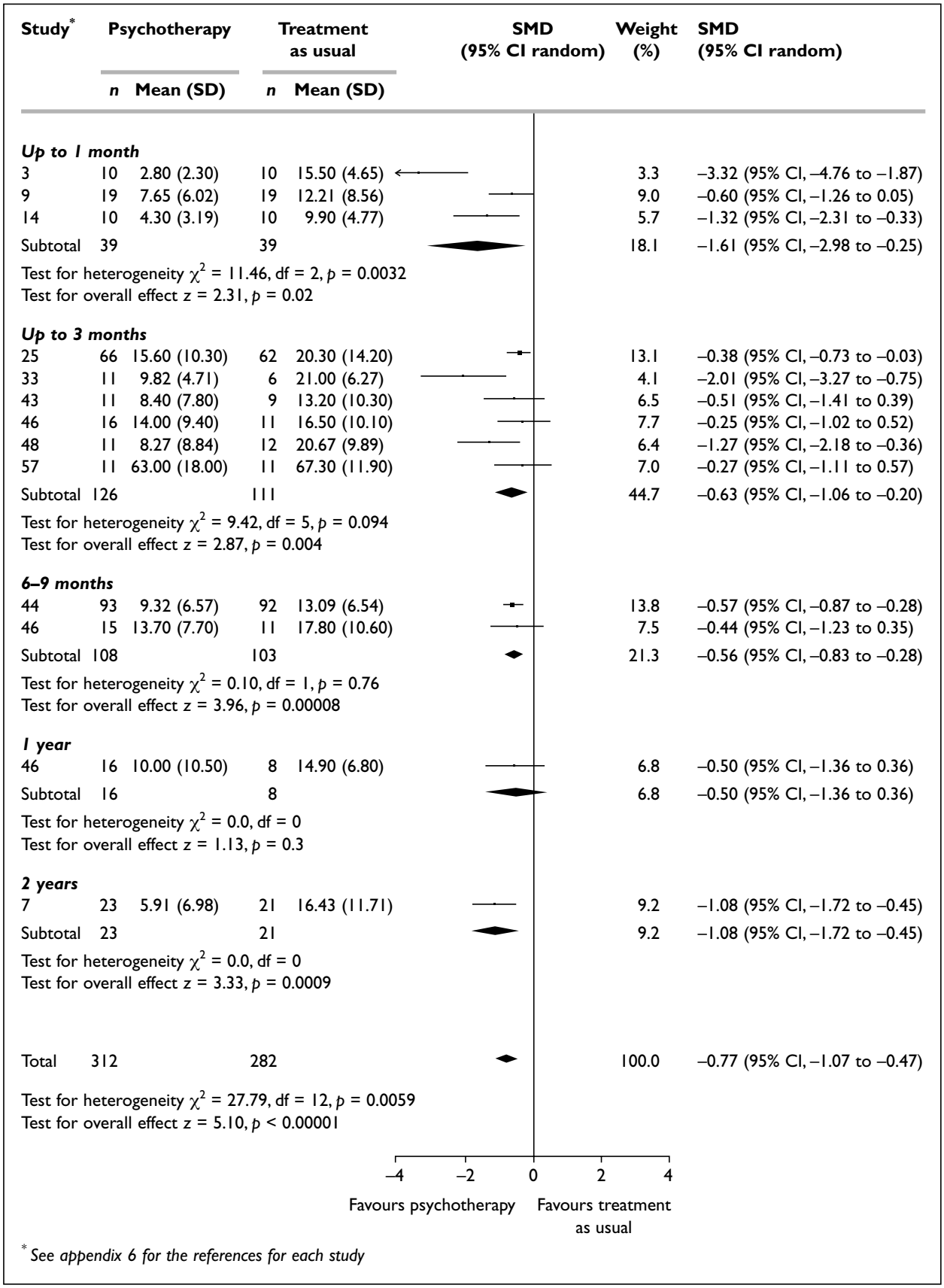


FIGURE 5 *Number of patients who were non-symptomatic at follow-up for all variants of psychotherapy versus treatment as usual/a waiting-list control*



* *See appendix 6 for the references for each study*

FIGURE 6 *Mean differences at follow-up for all variants of psychotherapy versus treatment as usual/a waiting-list control*

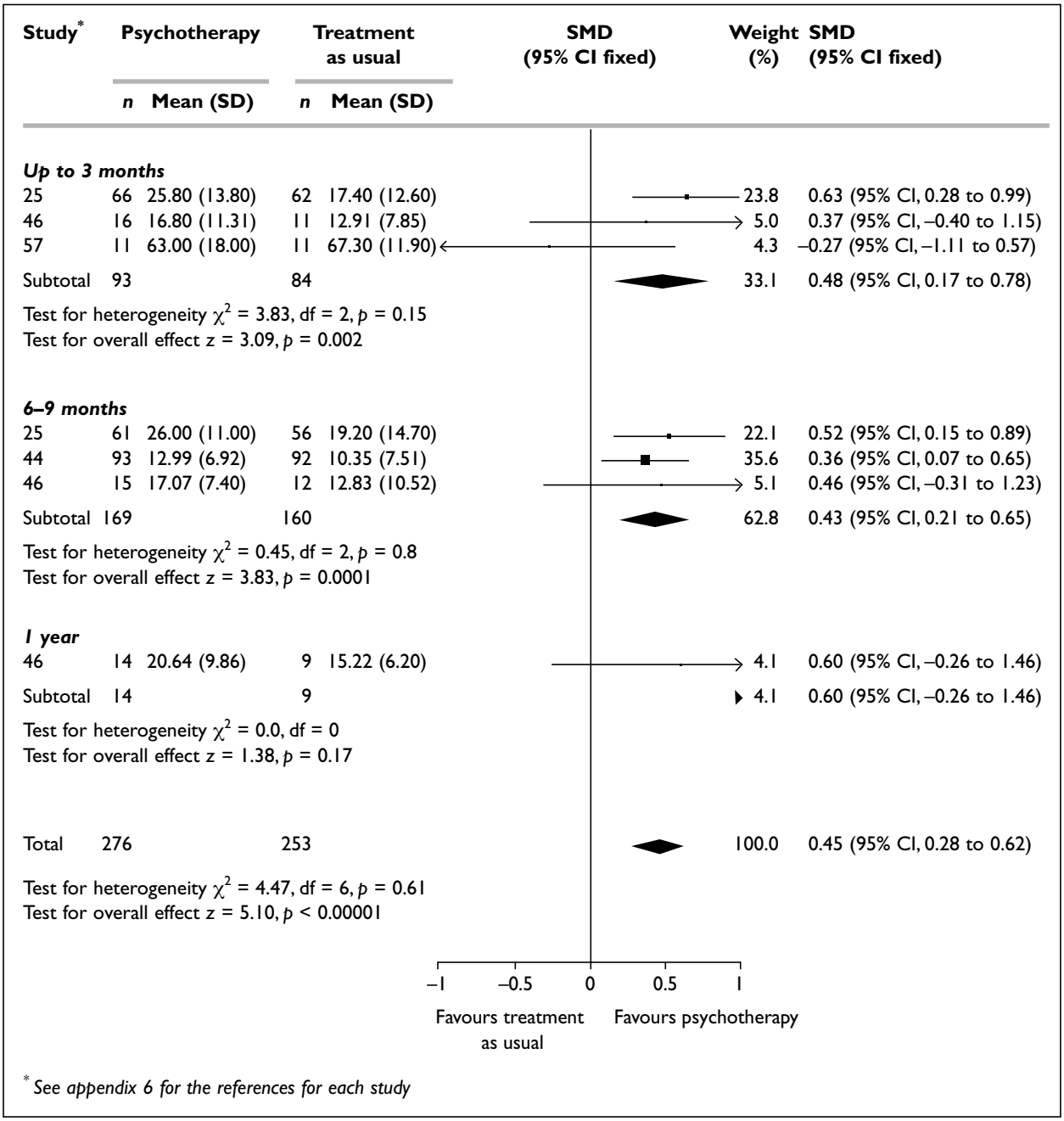
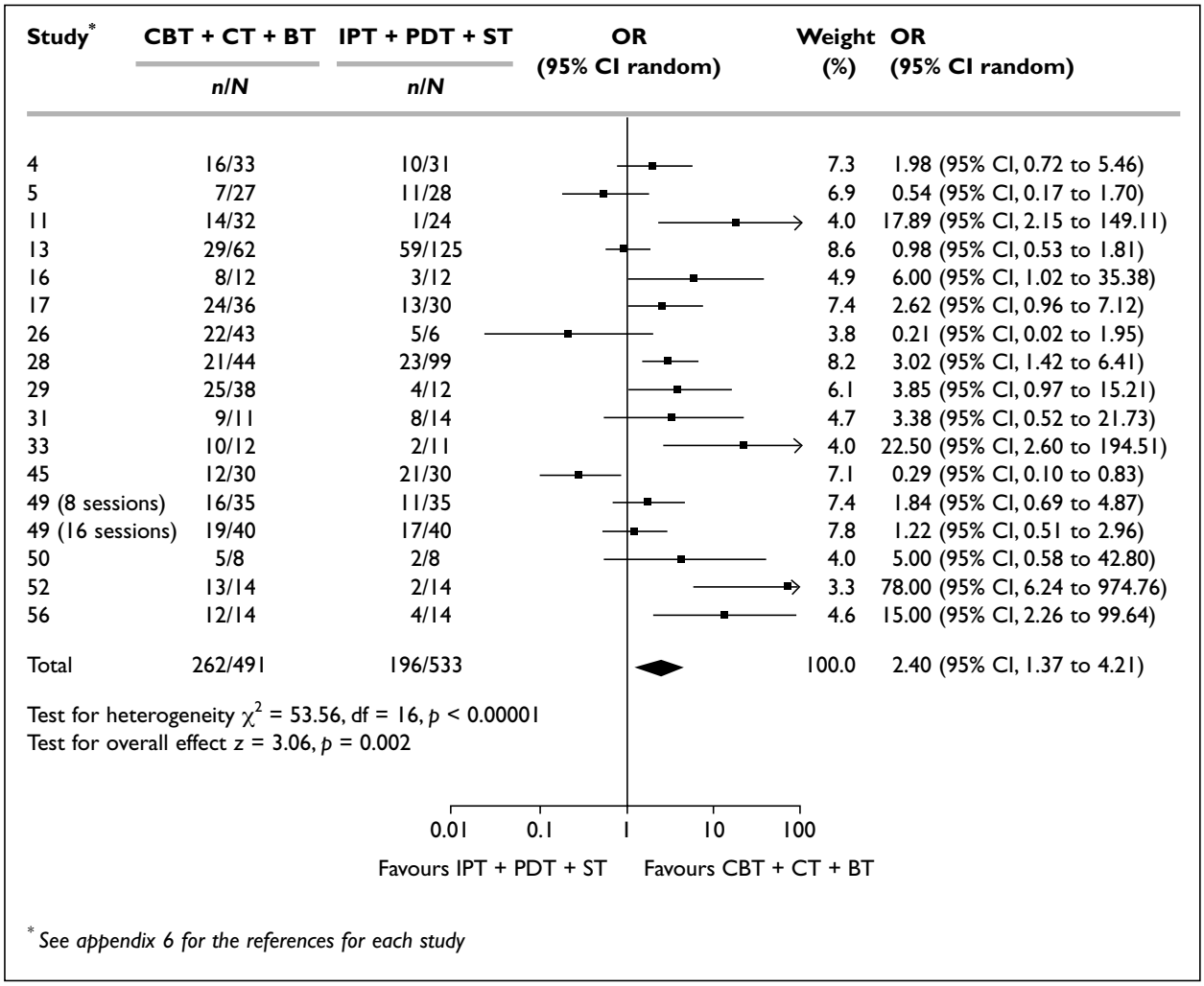


FIGURE 7 *Mean change from baseline at follow-up for all variants of psychotherapy versus treatment as usual/a waiting-list control*



* *See appendix 6 for the references for each study*

FIGURE 8 *Recovery at post-treatment for CBT + CT + BT versus PDT + IPT + ST*

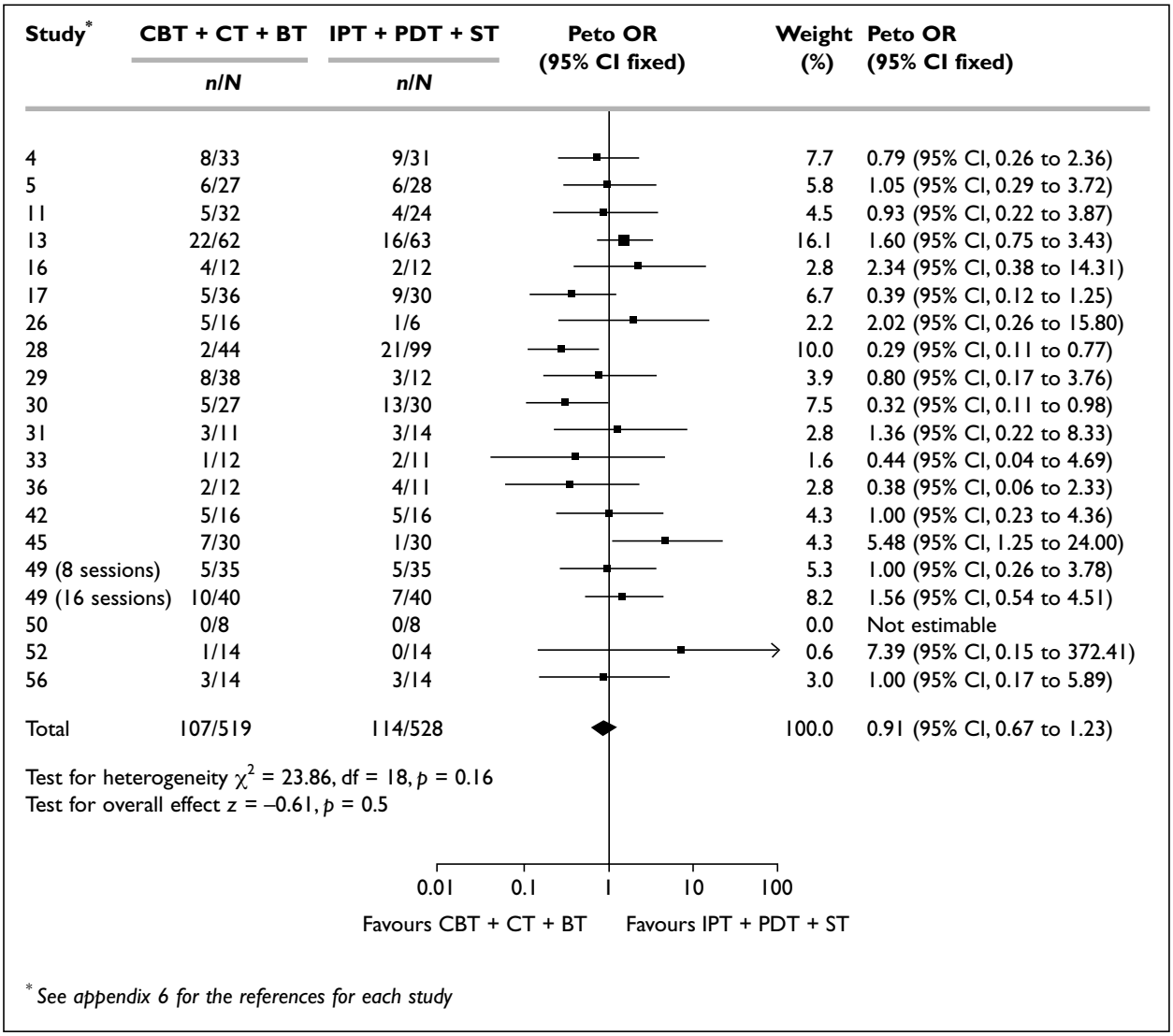
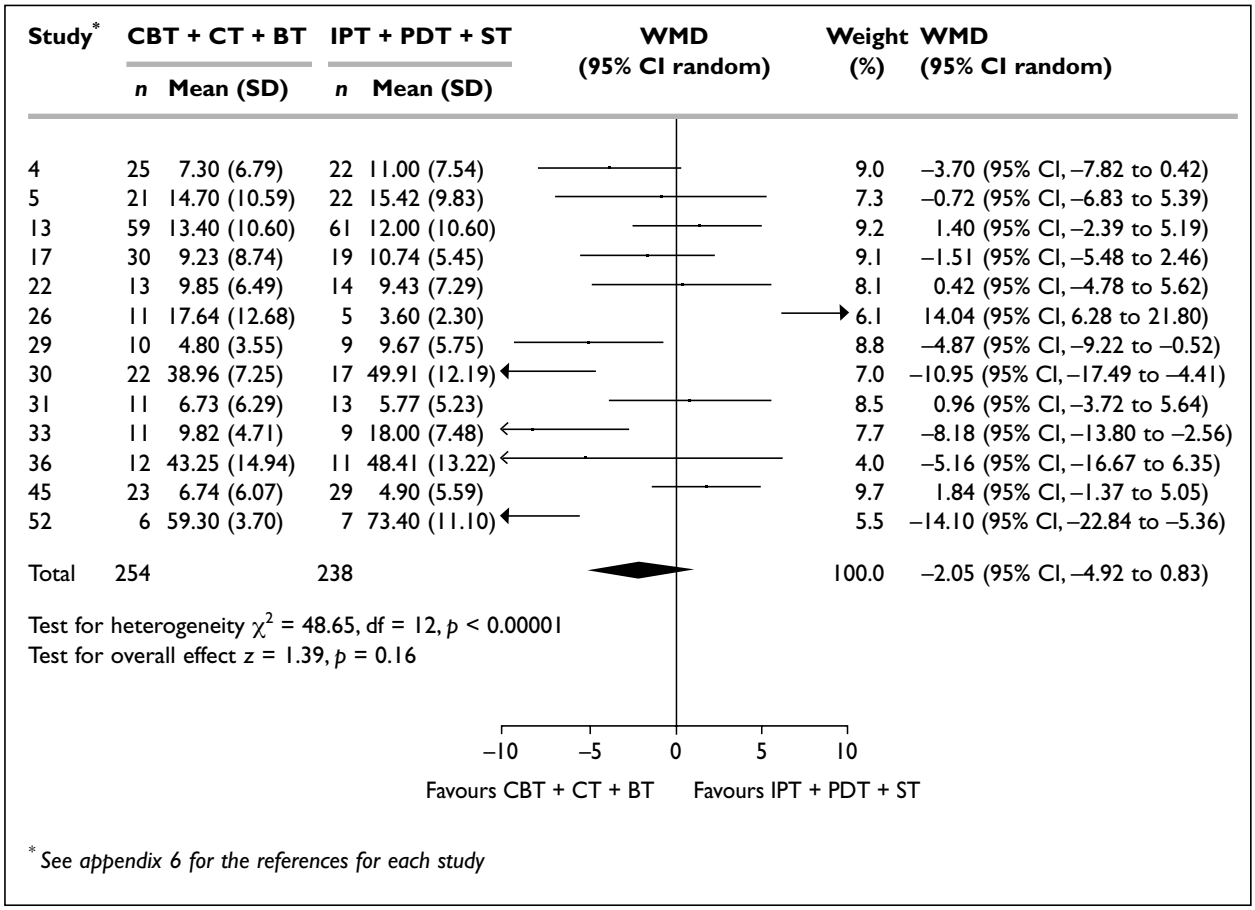
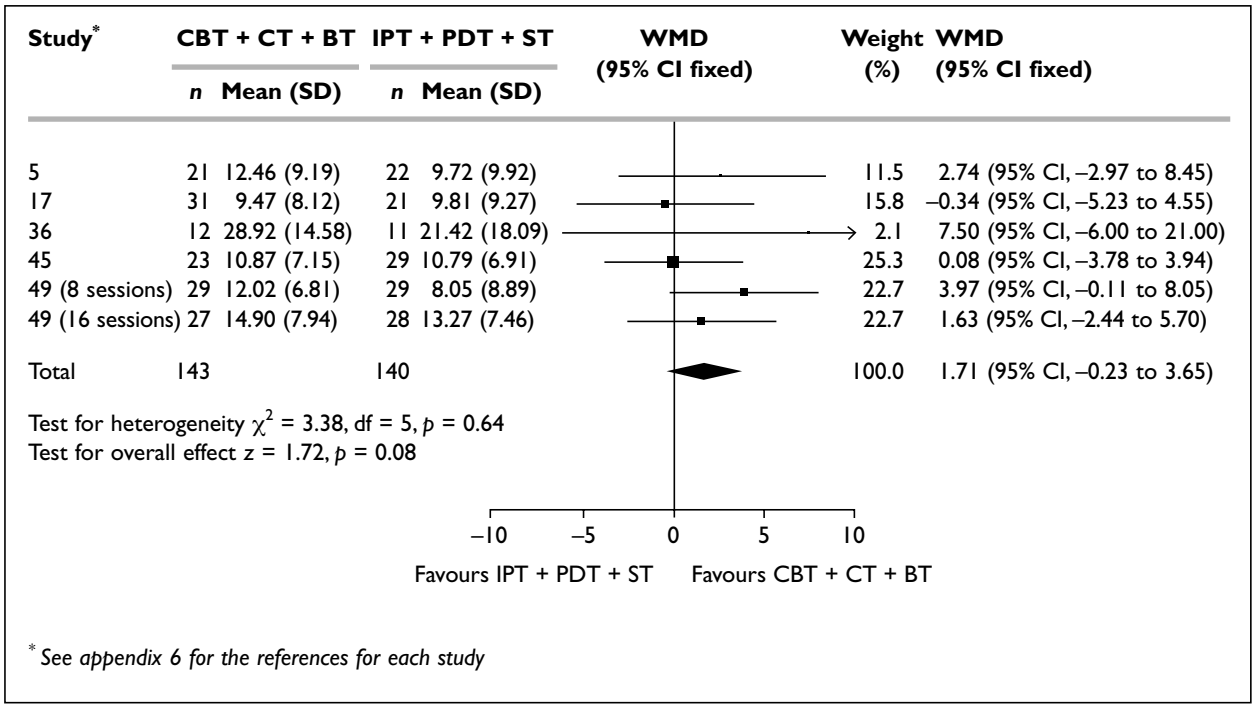


FIGURE 9 *Reported dropouts at post-treatment for CBT + CT + BT versus PDT + IPT + ST*



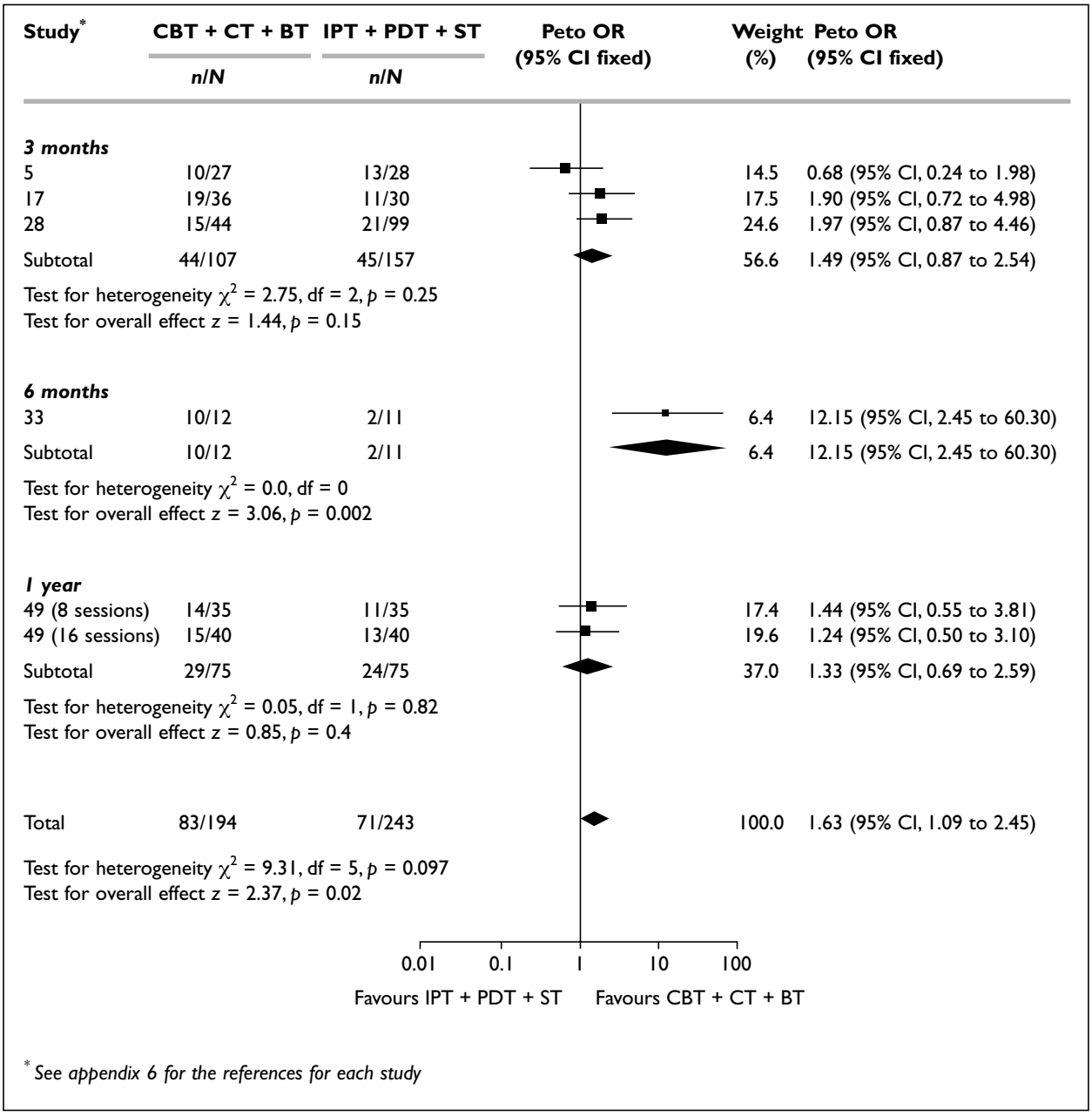
* *See appendix 6 for the references for each study*

FIGURE 10 *Mean differences post-treatment for CBT + CT + BT versus PDT + IPT + ST*



* *See appendix 6 for the references for each study*

FIGURE 11 *Mean change from baseline at post-treatment for CBT + CT + BT versus PDT + IPT + ST*



* *See appendix 6 for the references for each study*

FIGURE 12 *Number of patients who were non-symptomatic at follow-up for CBT + CT + BT versus PDT + IPT + ST*

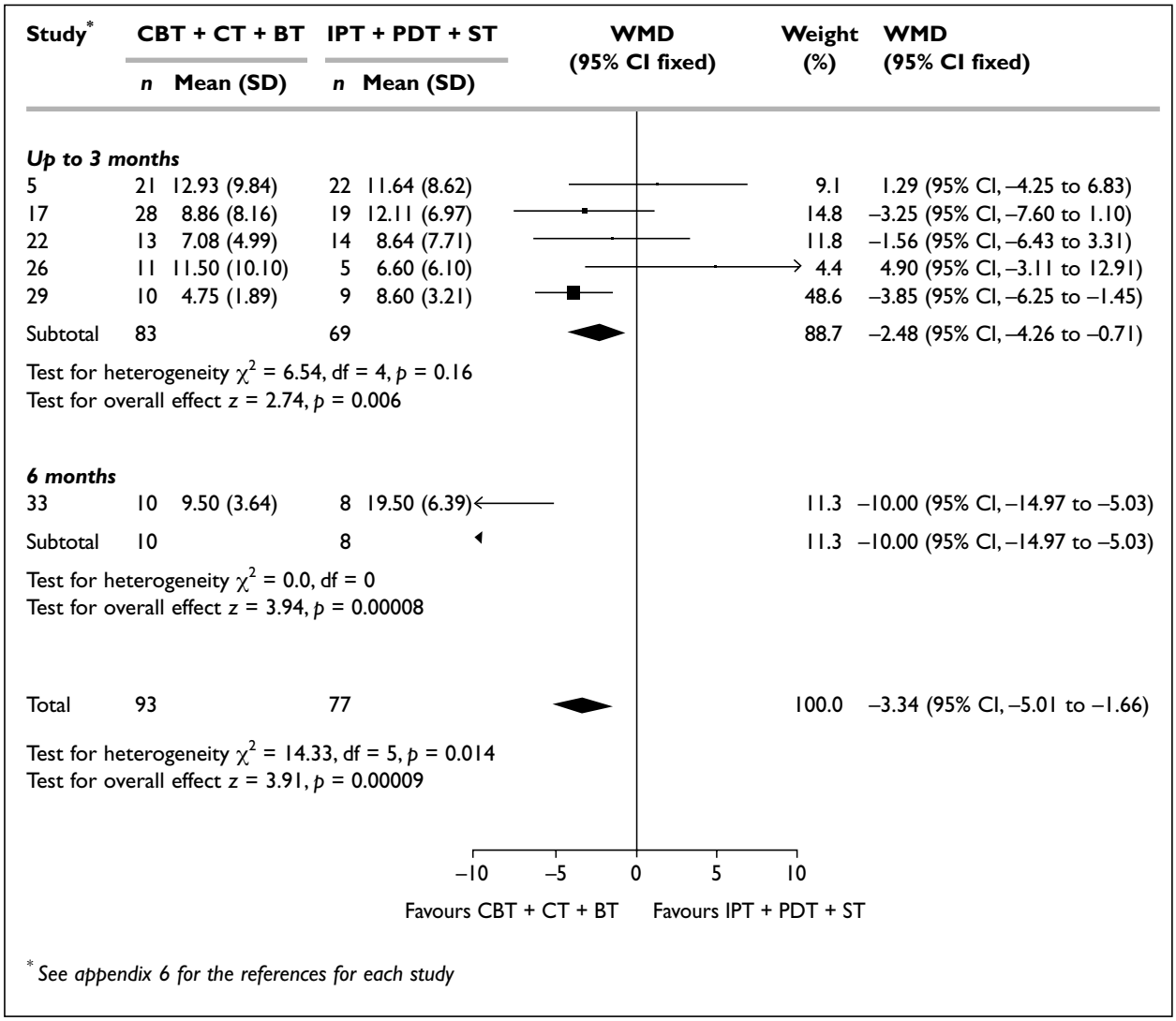


FIGURE 13 *Mean differences at follow-up for CBT + CT + BT versus PDT + IPT + ST*

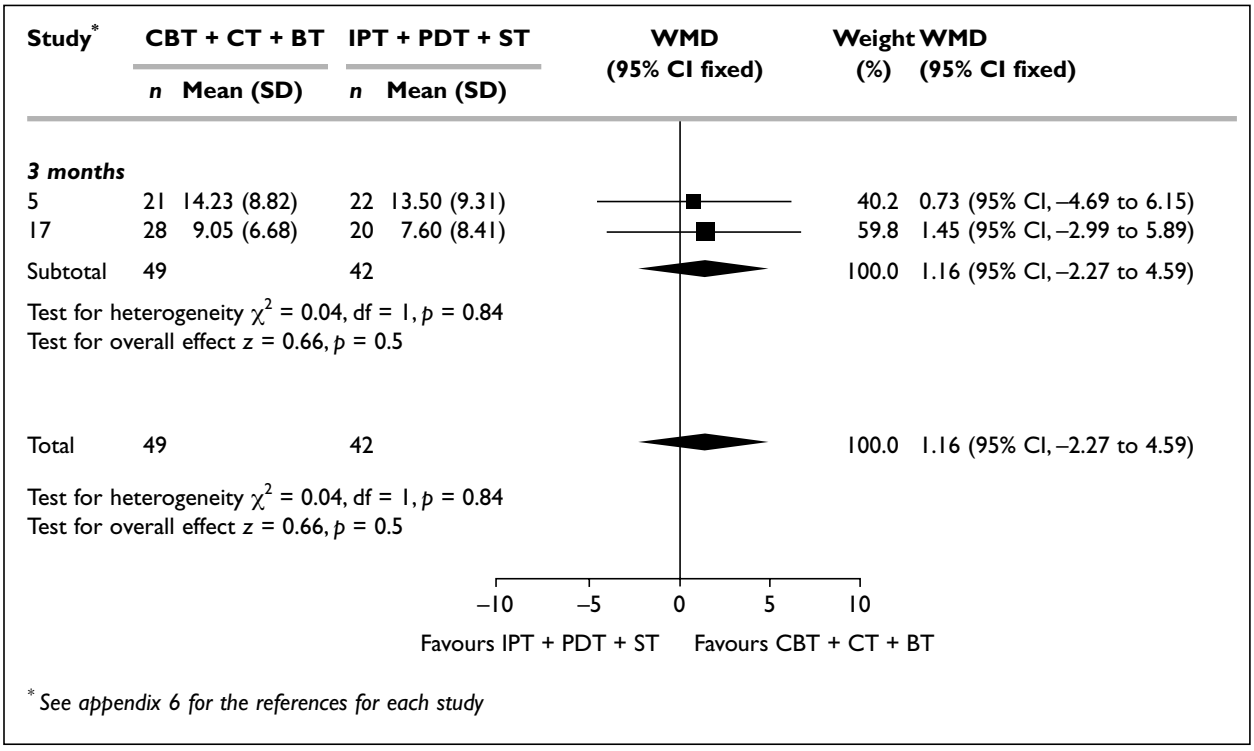


FIGURE 14 *Mean change from baseline at follow-up for CBT + CT + BT versus PDT + IPT + ST*

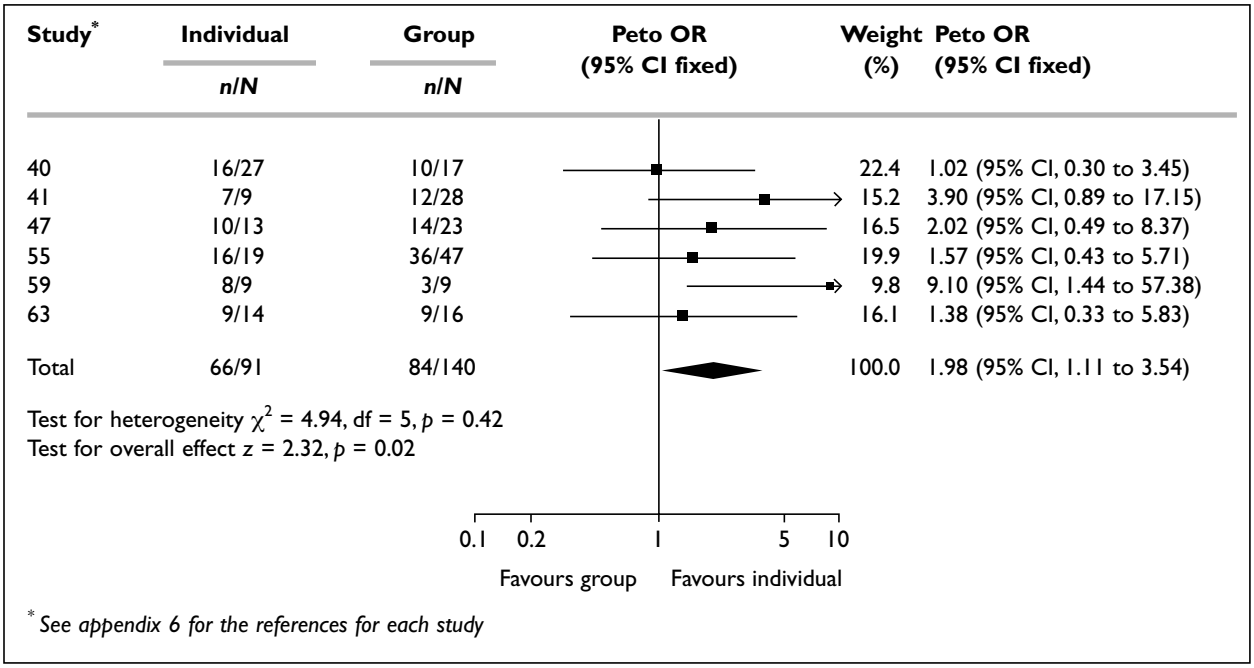


FIGURE 15 *Recovery at post-treatment for individual versus group therapy (all variants)*

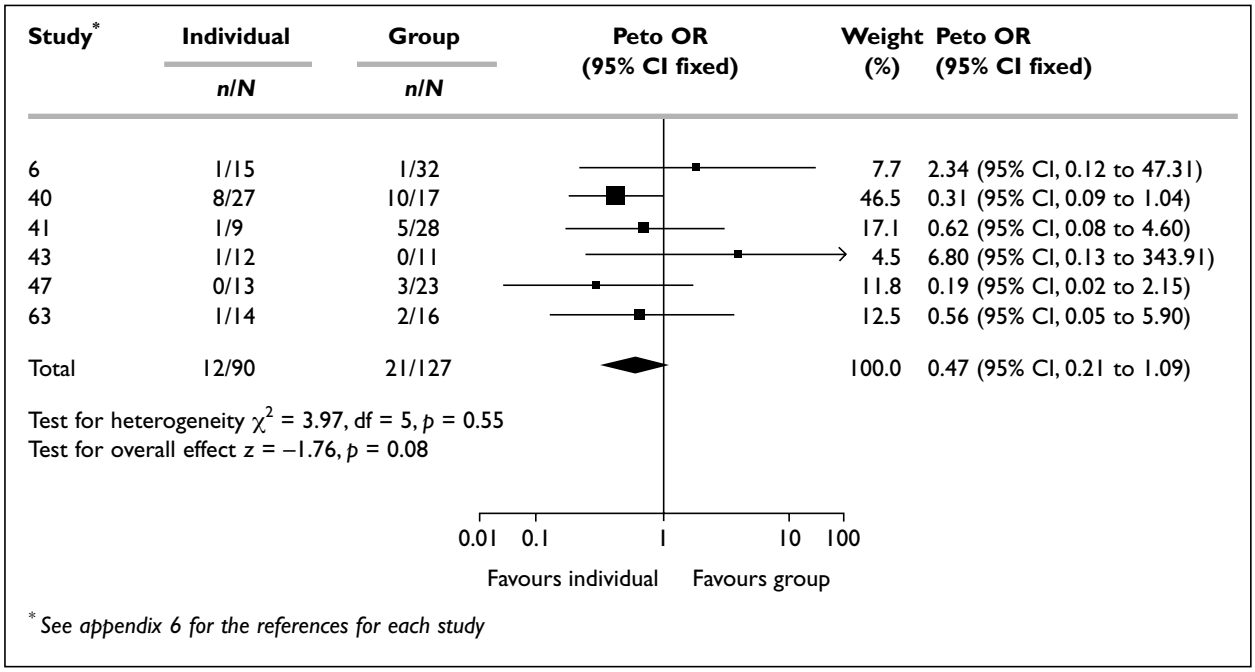


FIGURE 16 *Reported dropouts at post-treatment for individual versus group therapy (all variants)*

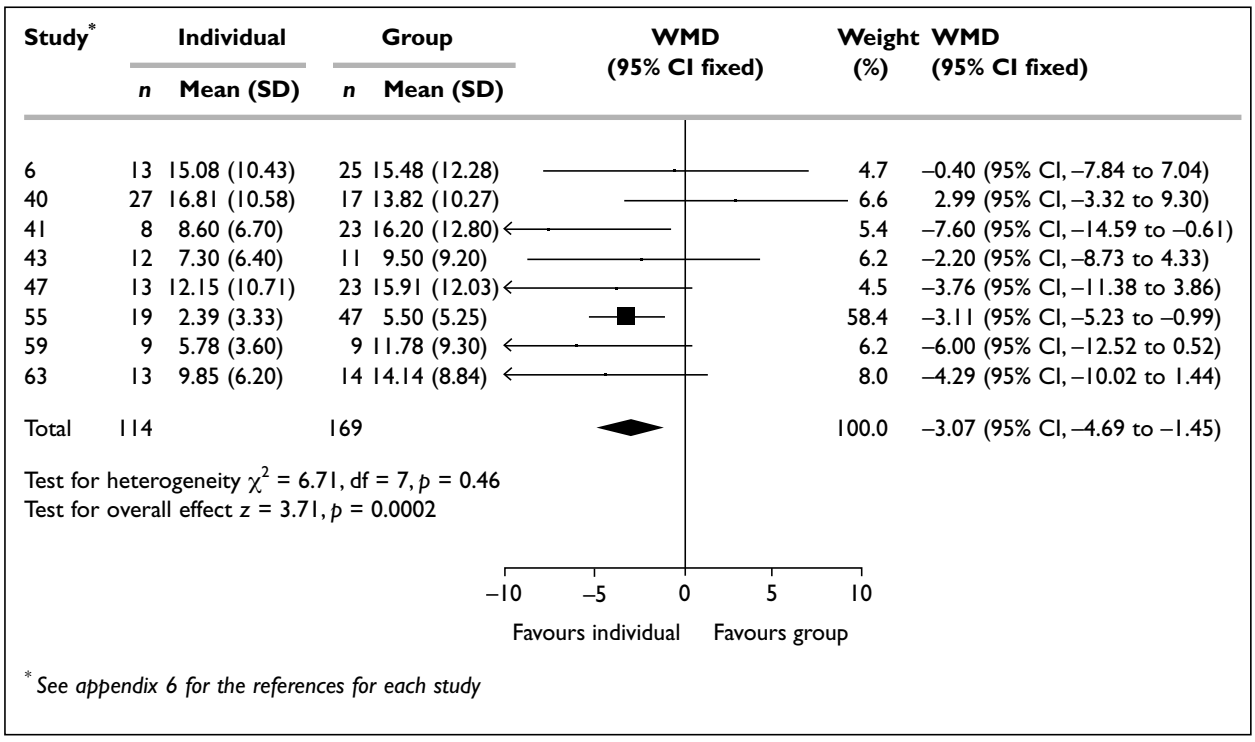


FIGURE 17 *Mean differences at post-treatment for individual versus group therapy (all variants)*

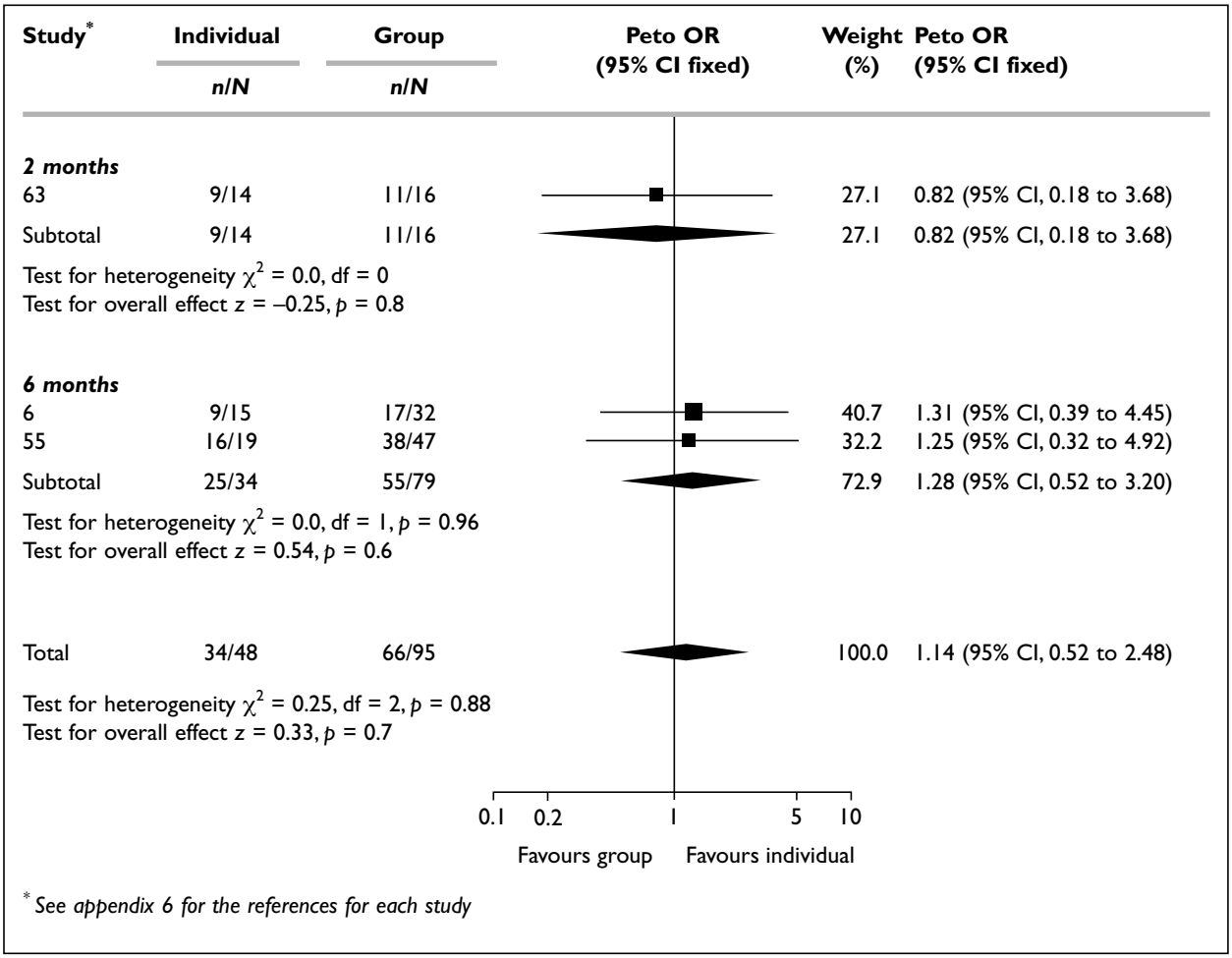


FIGURE 18 *Number of patients who were non-symptomatic at follow-up for individual versus group therapy (all variants)*

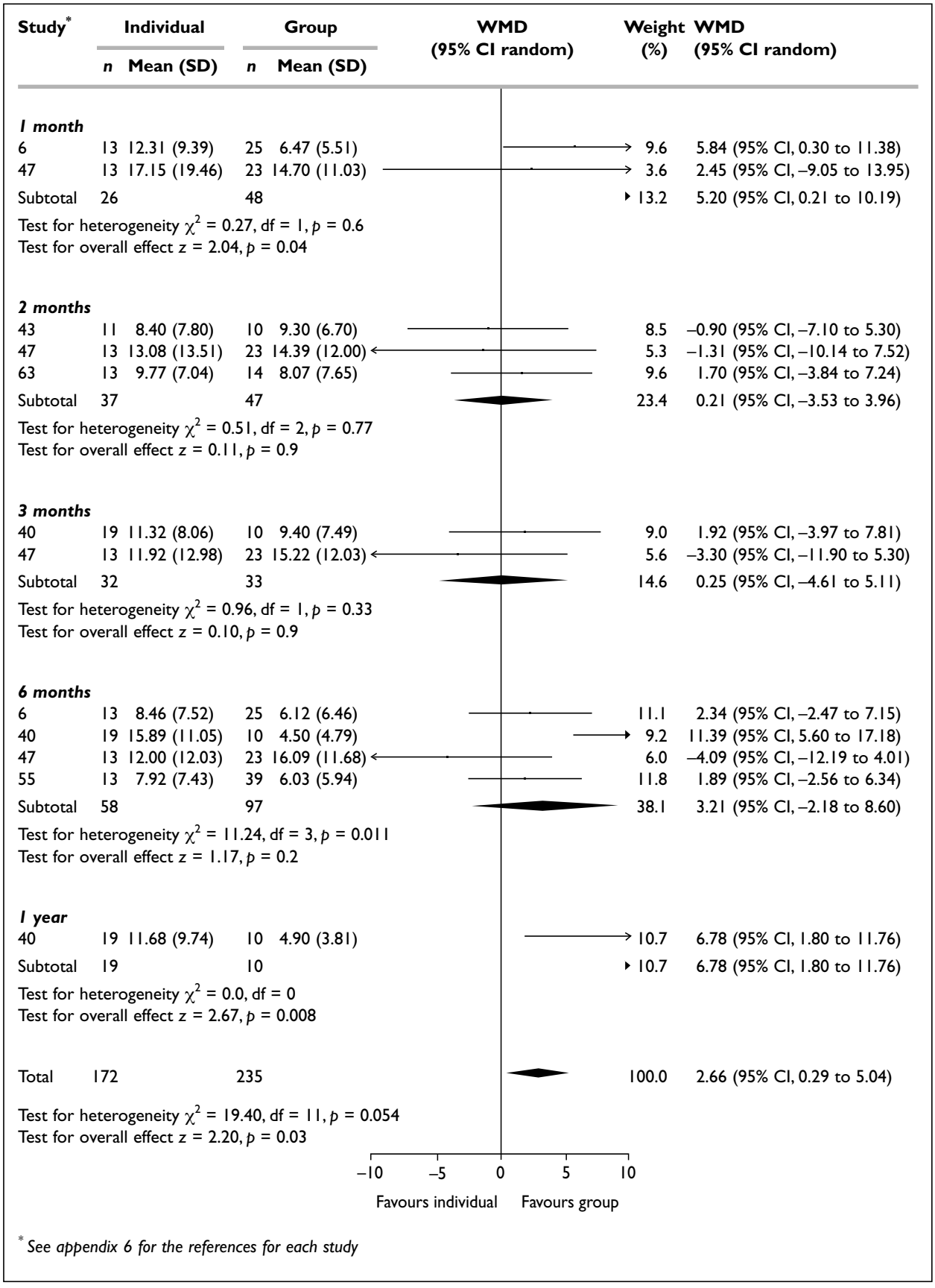


FIGURE 19 *Mean differences at follow-up for individual versus group therapy (all variants)*

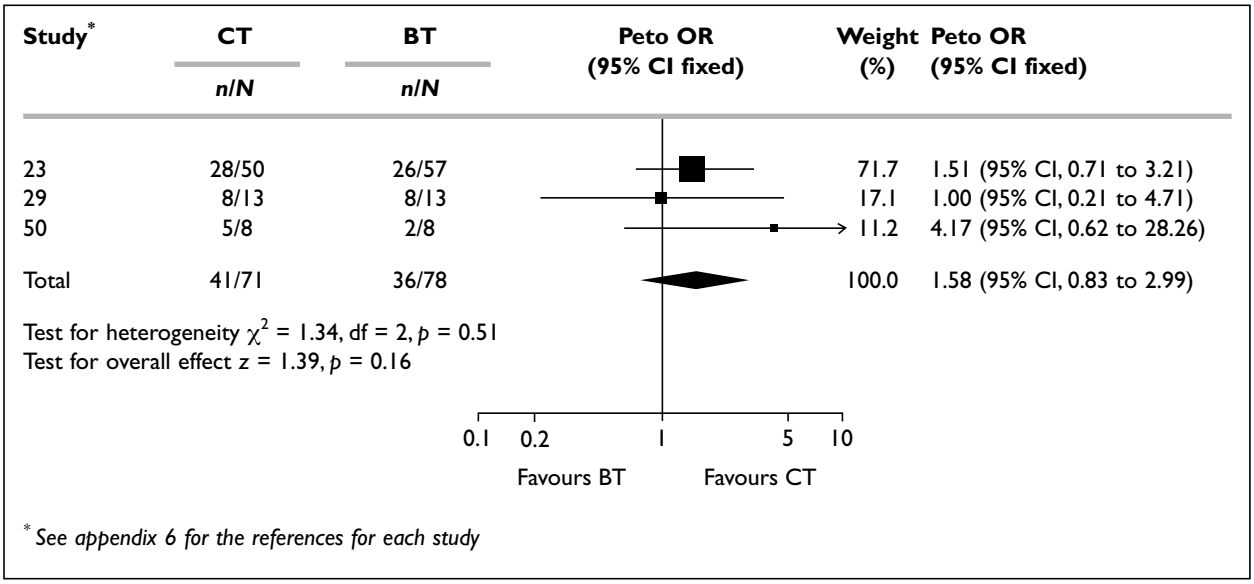


FIGURE 20 *Recovery at post-treatment for CT versus BT*

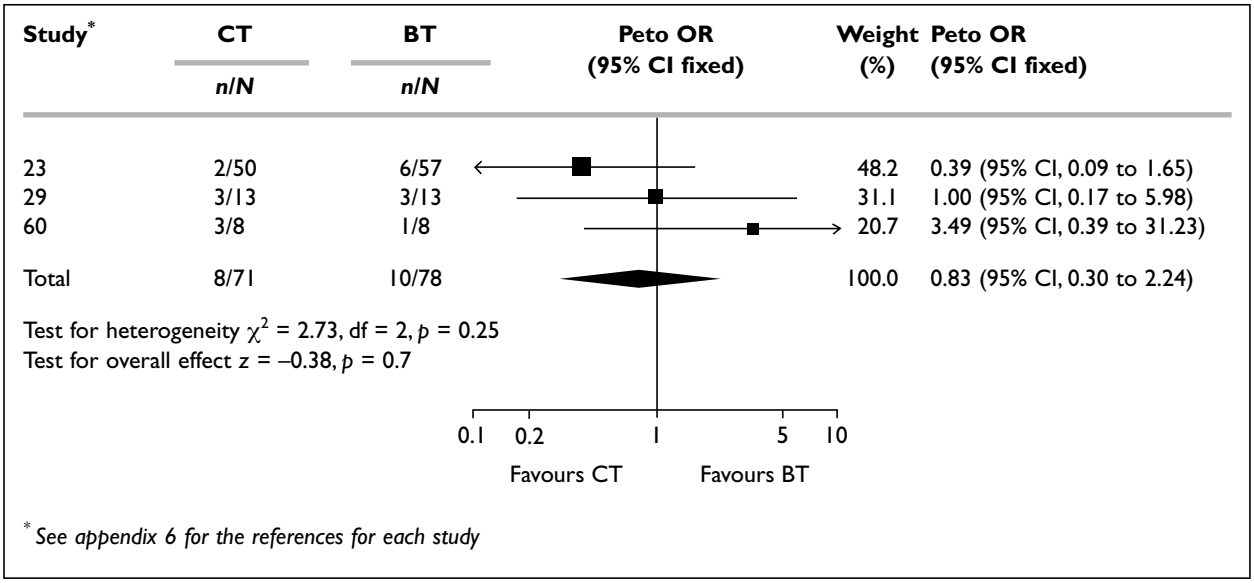


FIGURE 21 *Reported dropouts at post-treatment for CT versus BT*

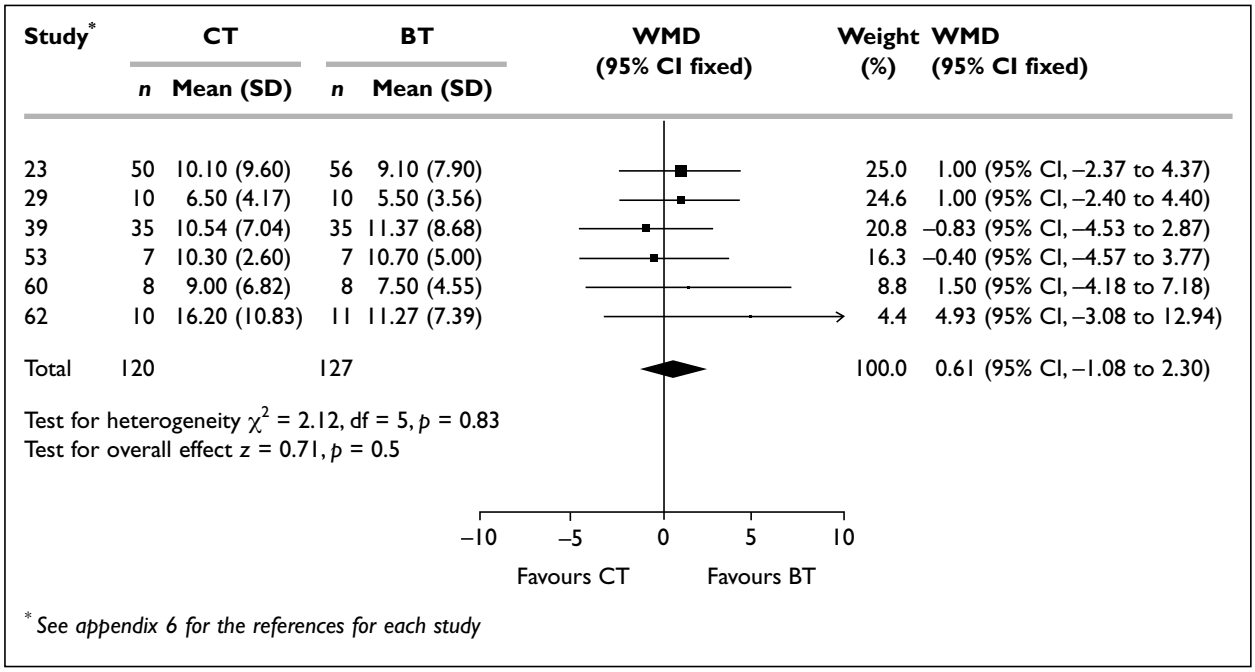


FIGURE 22 *Mean differences at post-treatment for CT versus BT*

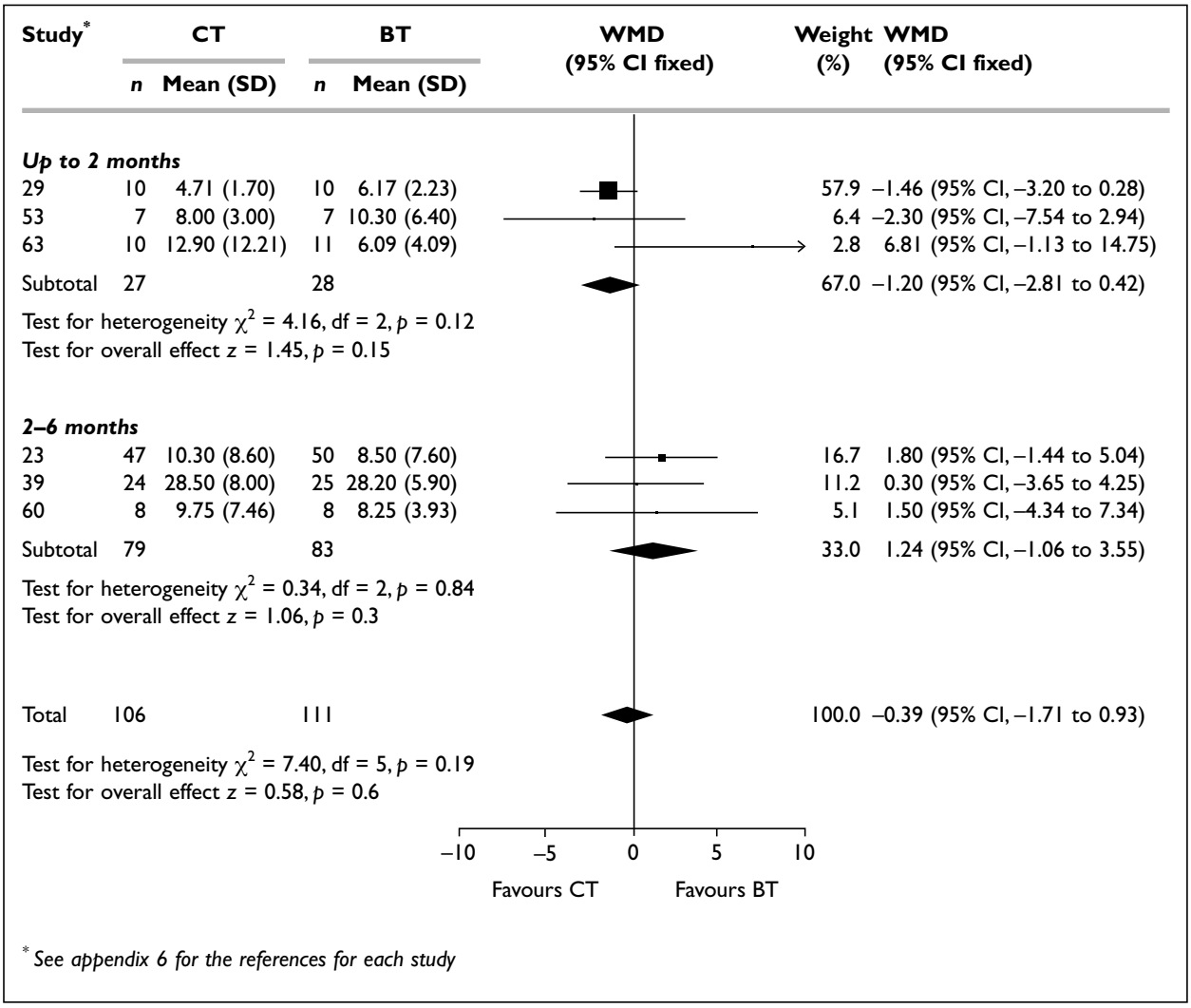


FIGURE 23 *Mean differences at follow-up for CT versus BT*

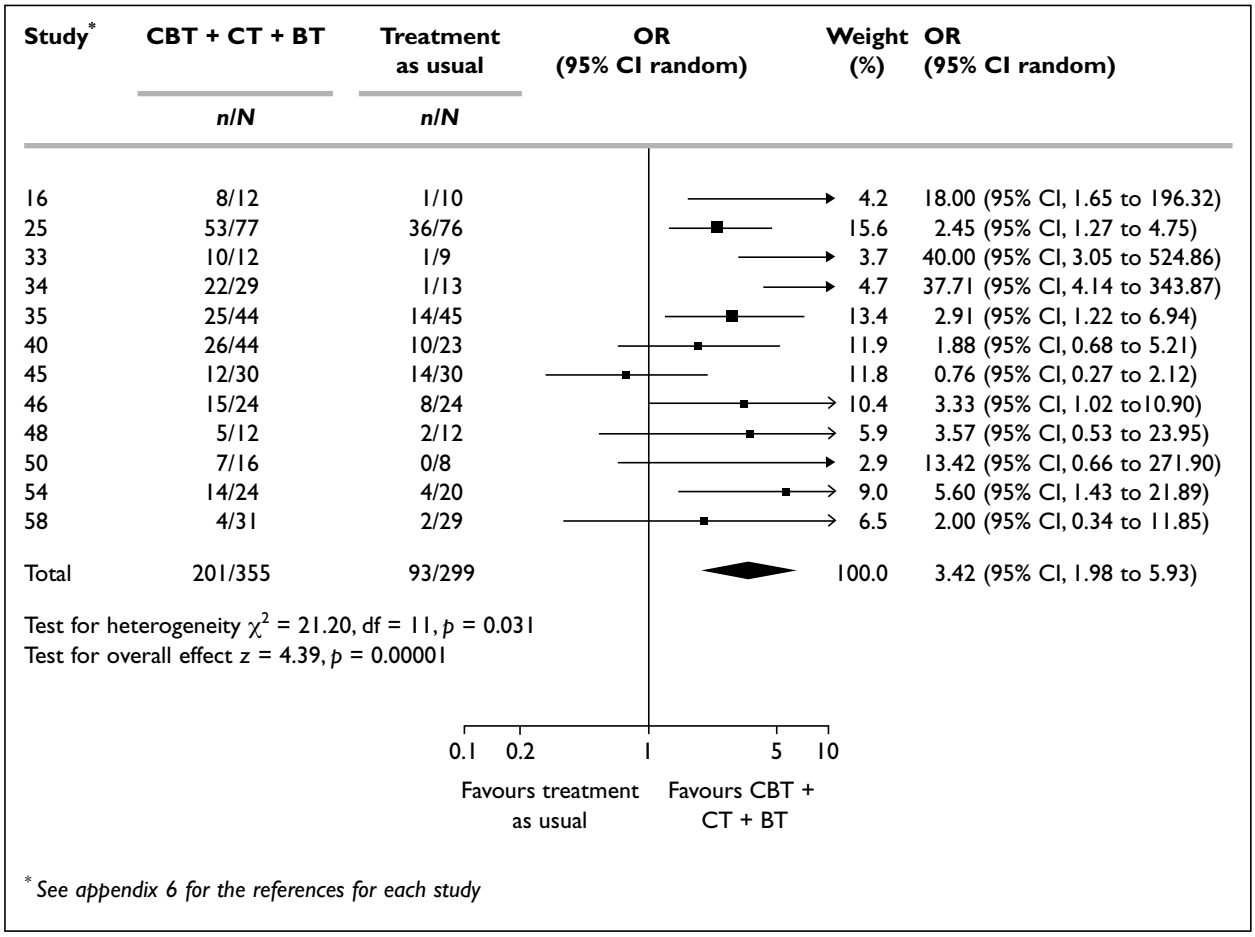


FIGURE 24 *Recovery at post-treatment for CBT + CT + BT versus treatment as usual/a waiting-list control*

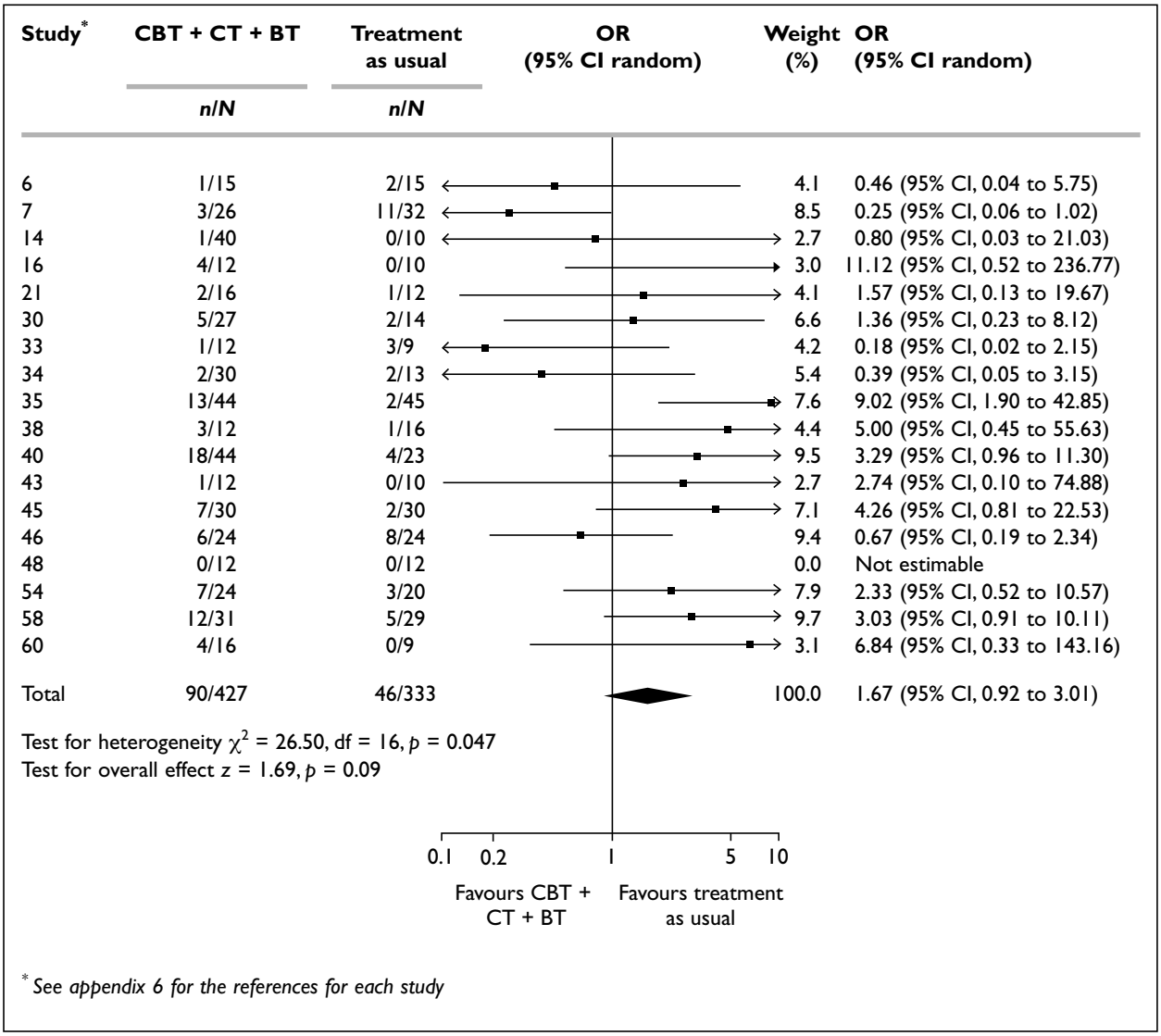
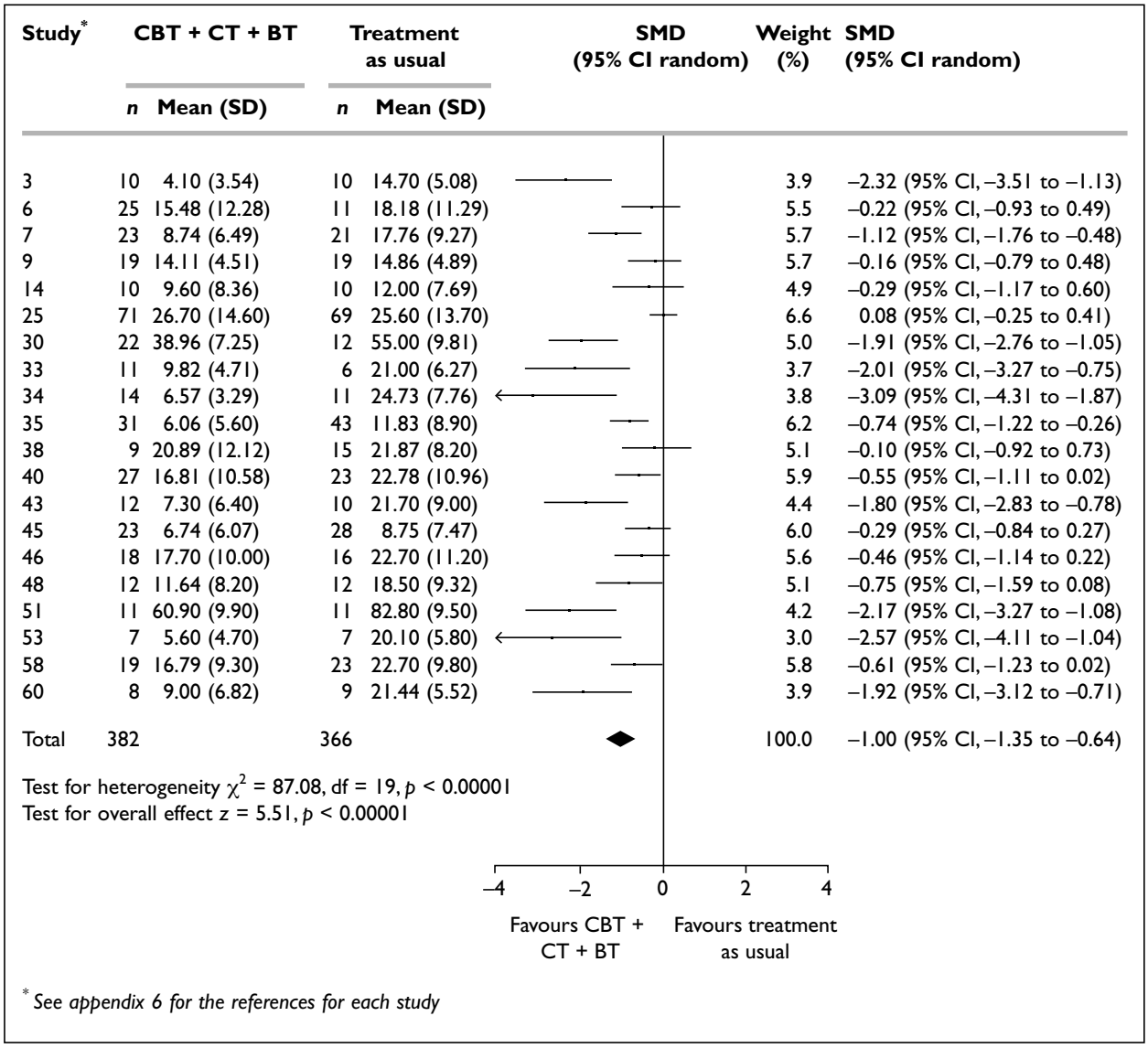


FIGURE 25 *Reported dropouts at post-treatment for CBT + CT + BT versus treatment as usual/a waiting-list control*



* *See appendix 6 for the references for each study*

FIGURE 26 *Mean differences at post-treatment for CBT + CT + BT versus treatment as usual/a waiting-list control*

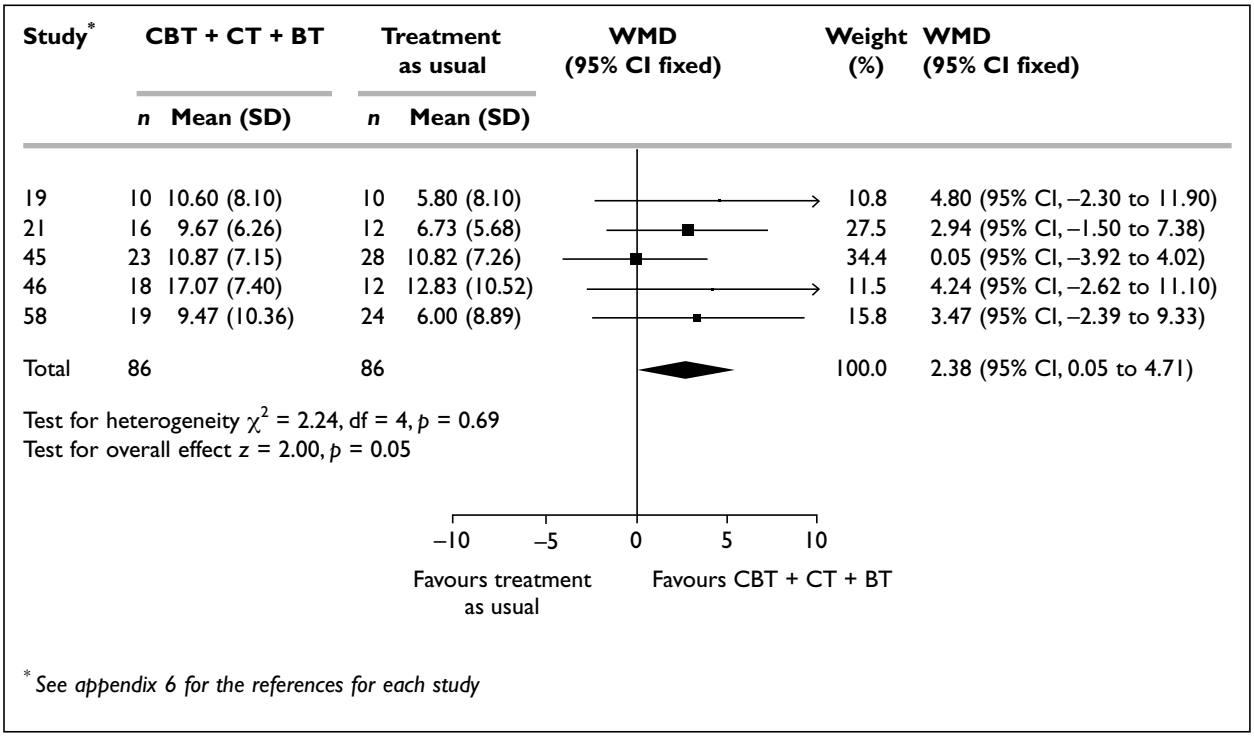


FIGURE 27 *Mean change from baseline at post-treatment for CBT + CT + BT versus treatment as usual/a waiting-list control*

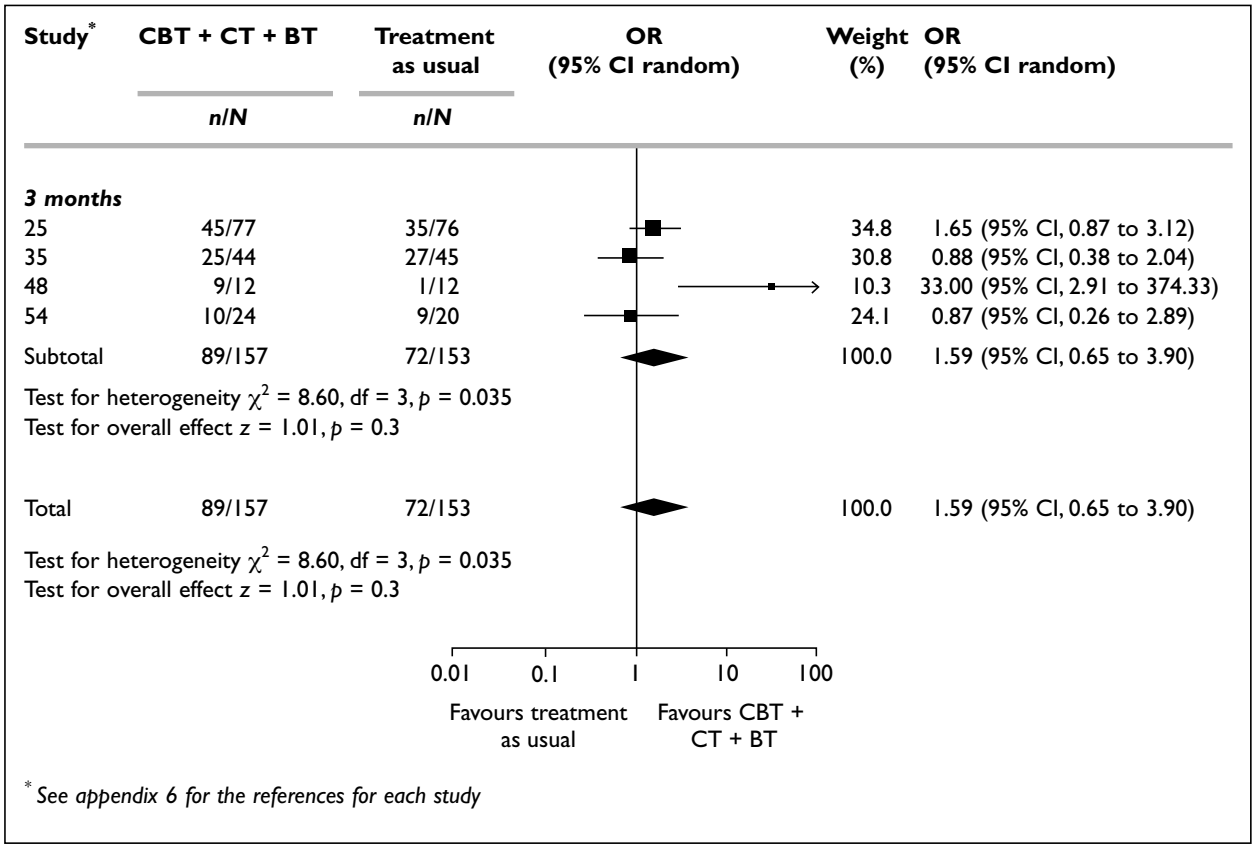
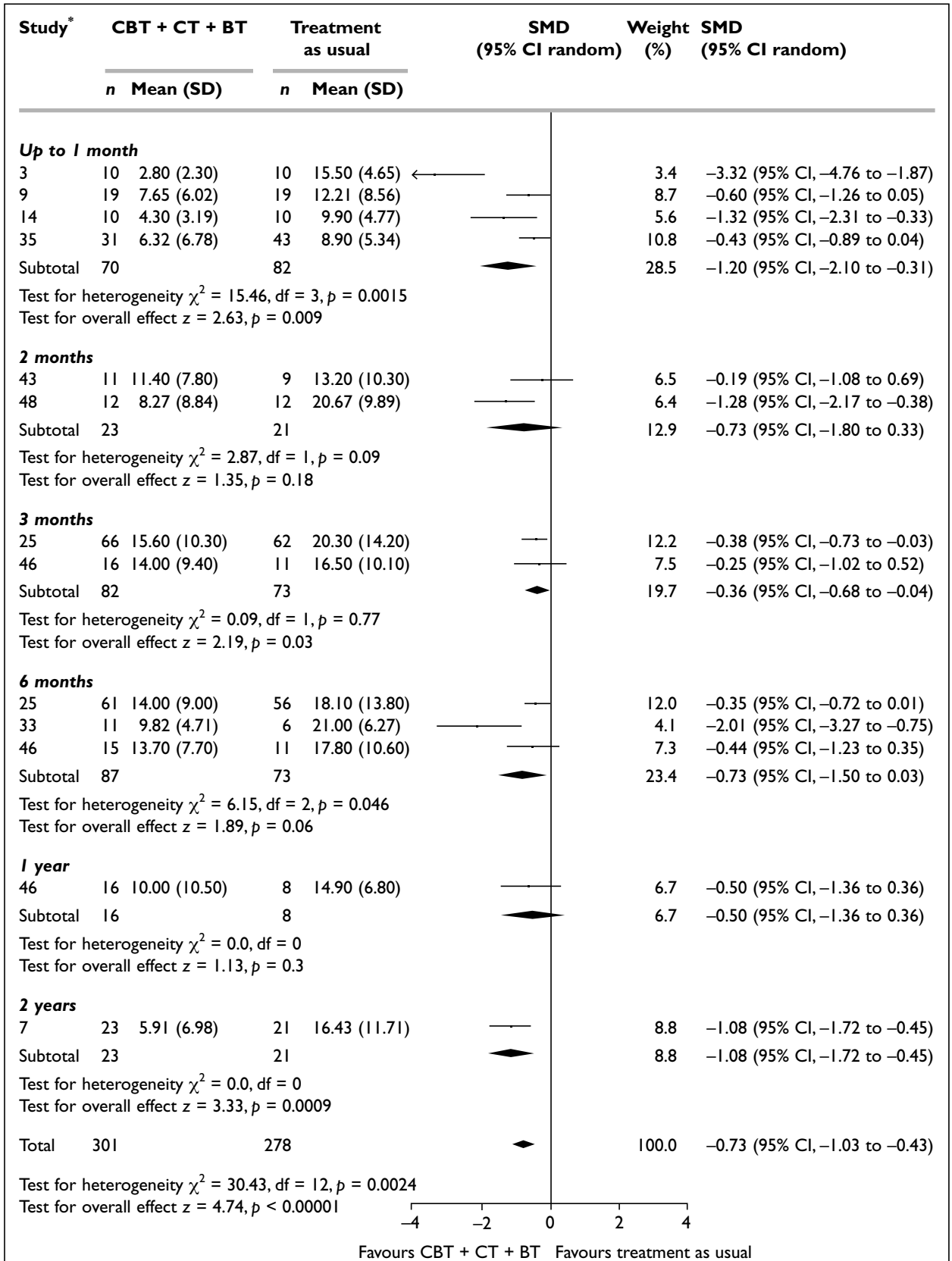


FIGURE 28 *Number of patients who were non-symptomatic at follow-up for CBT + CT + BT versus treatment as usual/a waiting-list control*



* *See appendix 6 for the references for each study*

FIGURE 29 *Mean differences at follow-up for CBT + CT + BT versus treatment as usual/a waiting-list control*

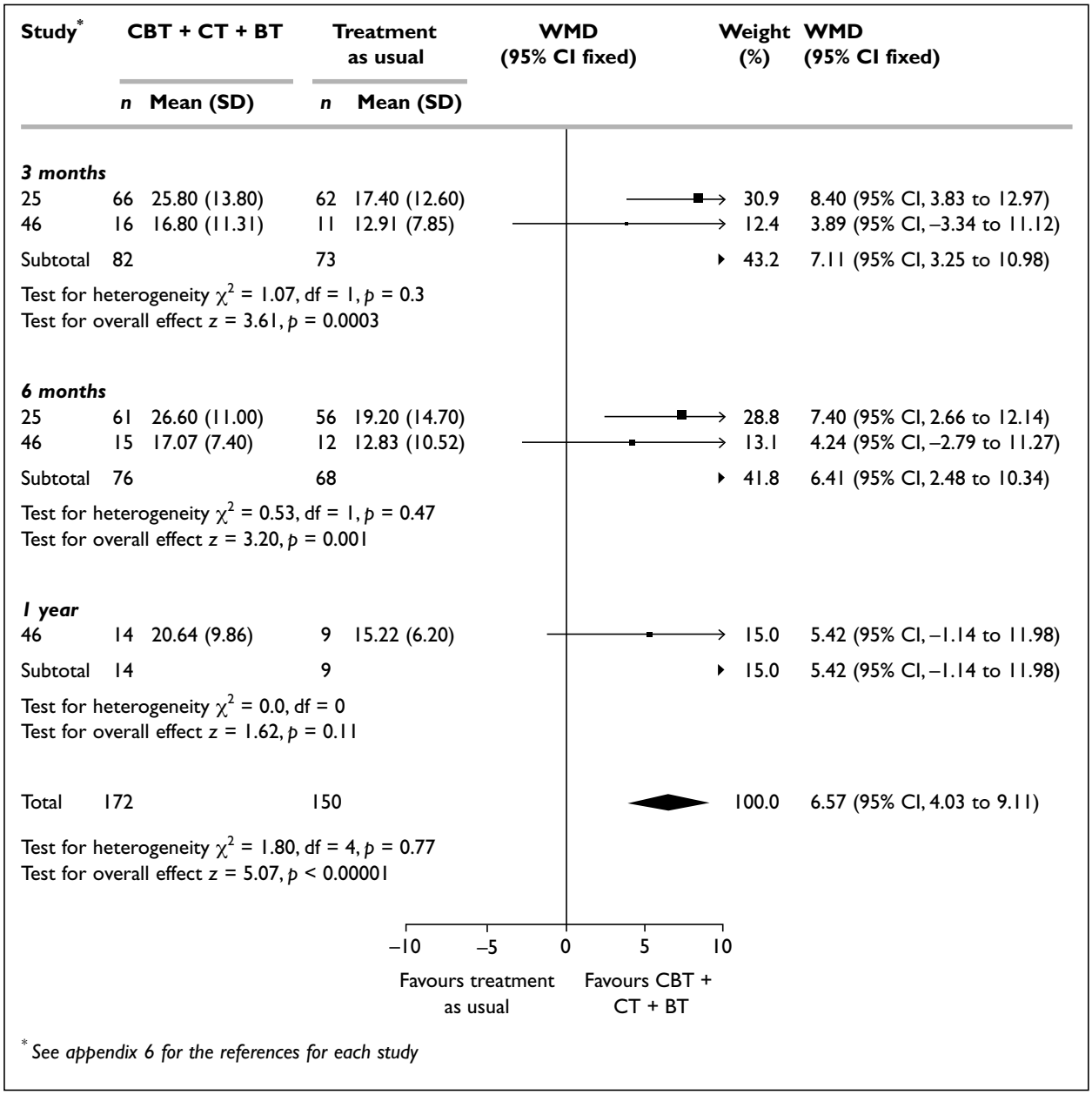


FIGURE 30 *Mean change from baseline at follow-up for CBT + CT + BT versus treatment as usual/a waiting-list control*

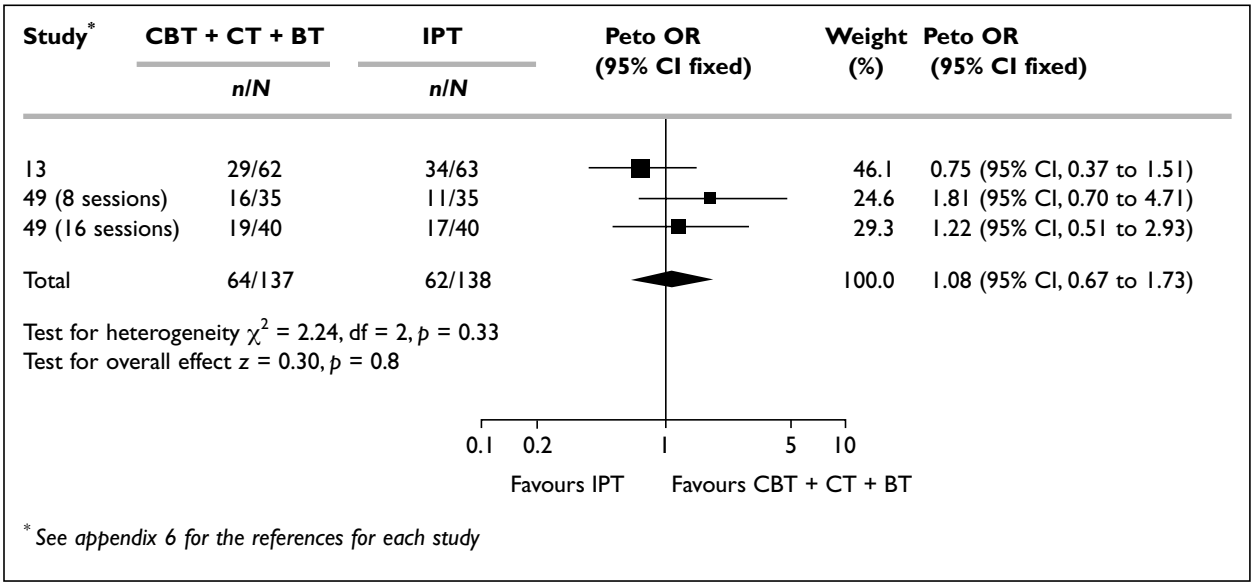


FIGURE 31 *Recovery at post-treatment for CBT + CT + BT versus IPT*

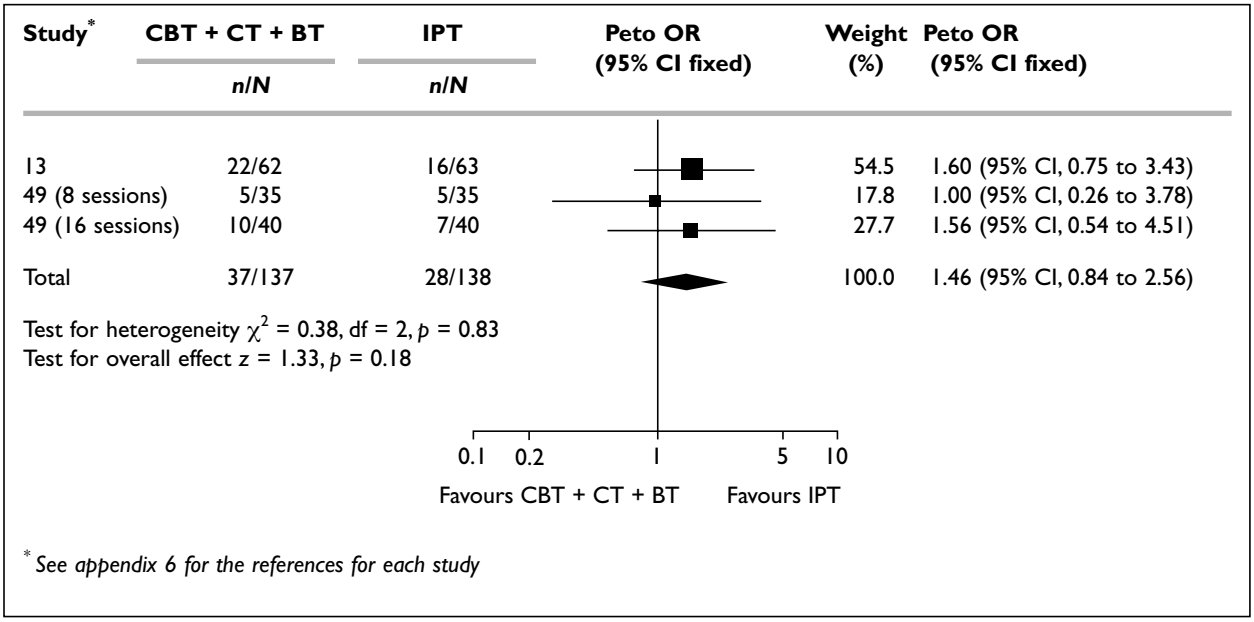


FIGURE 32 *Reported dropouts at post-treatment for CBT + CT + BT versus IPT*

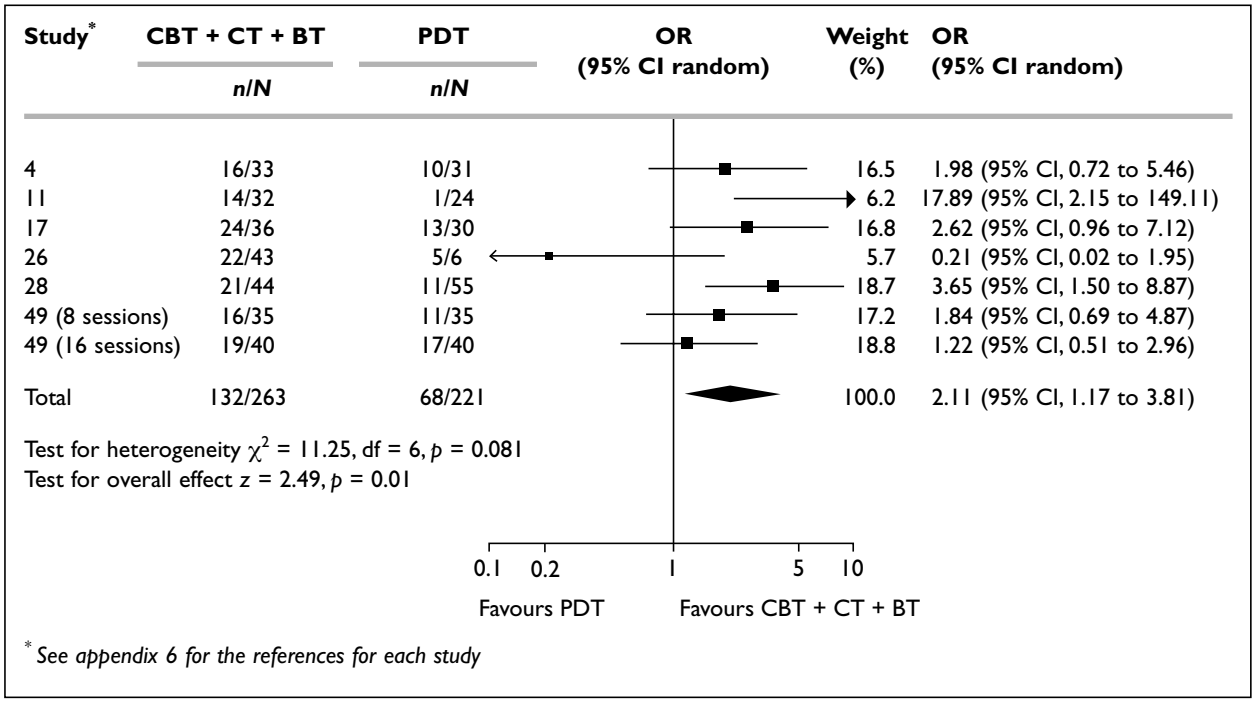


FIGURE 33 *Recovery at post-treatment for CBT + CT + BT versus PDT*

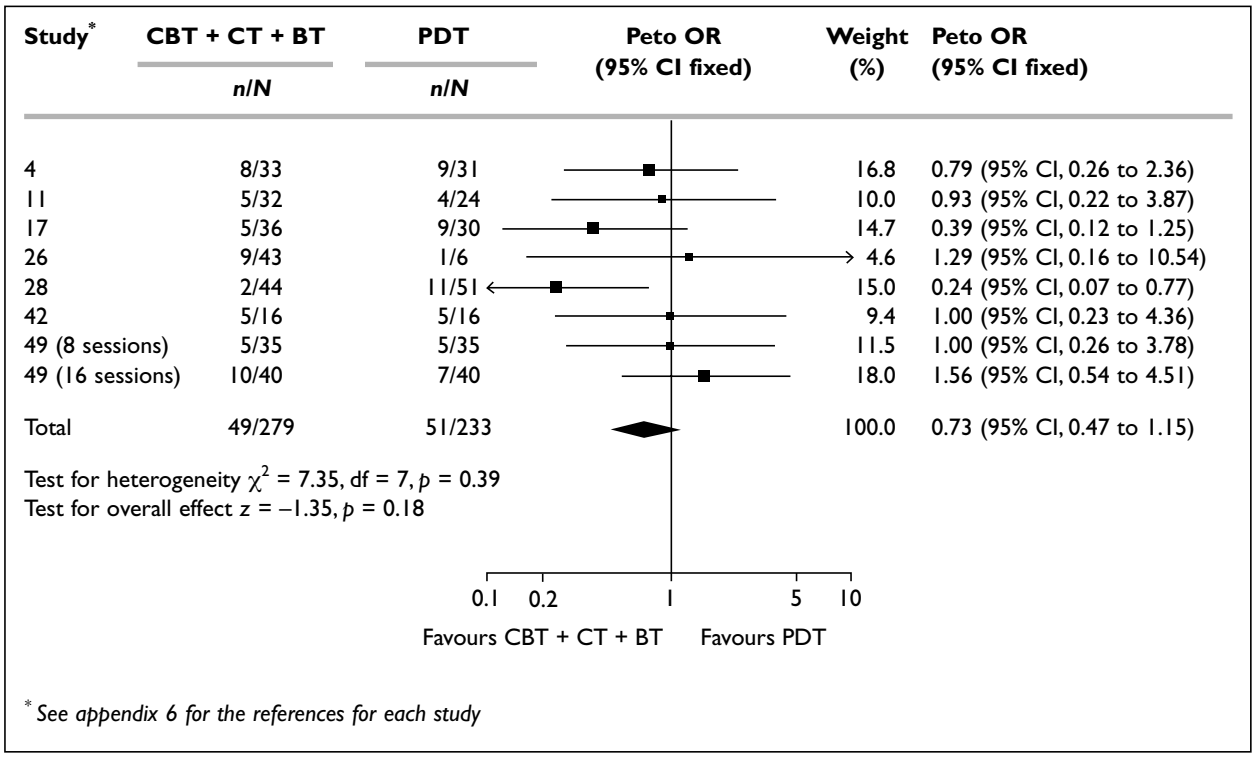


FIGURE 34 *Reported dropouts at post-treatment for CBT + CT + BT versus PDT*

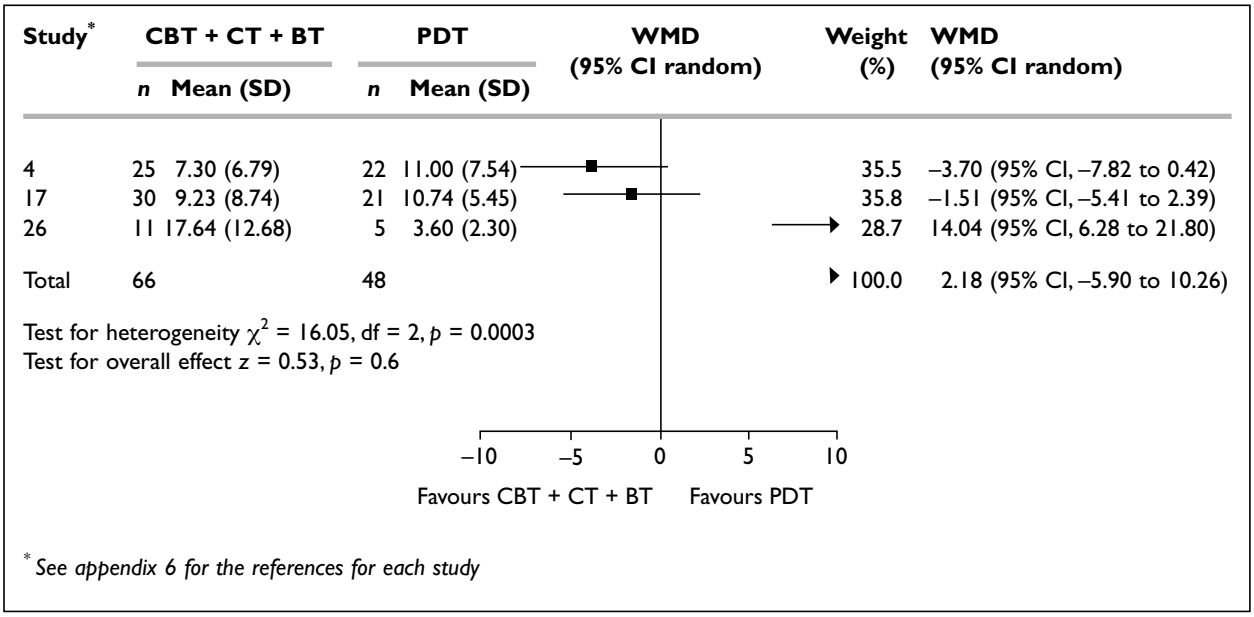


FIGURE 35 *Mean differences post-treatment for CBT + CT + BT versus PDT*

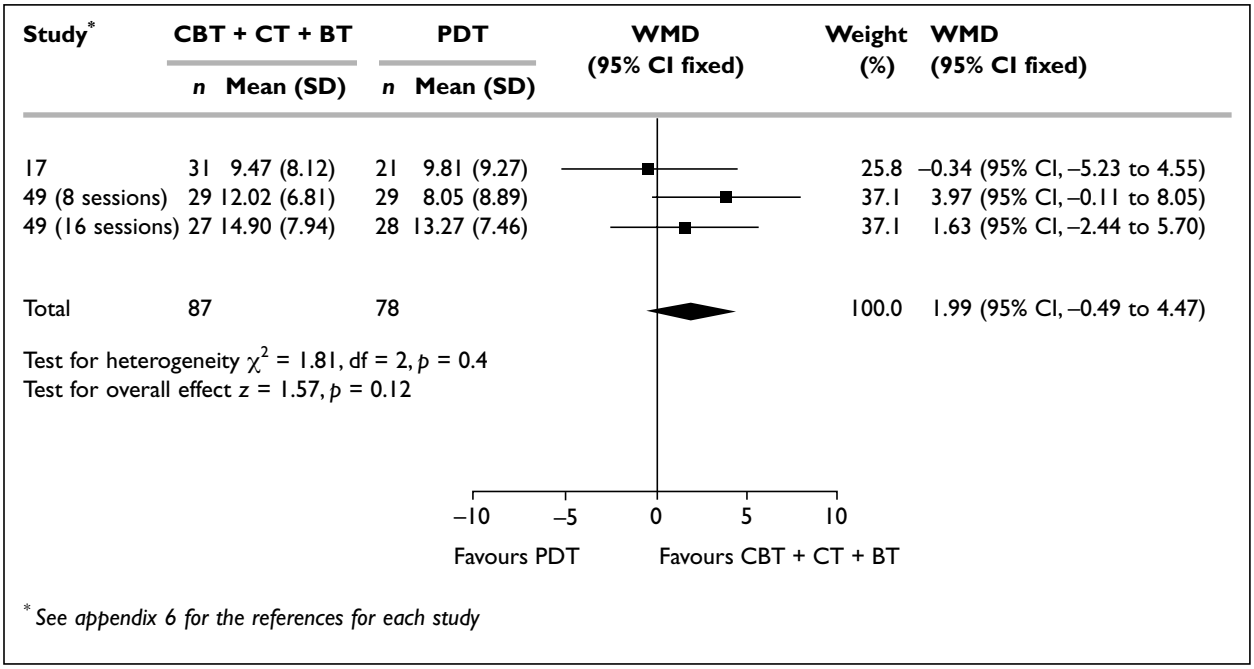
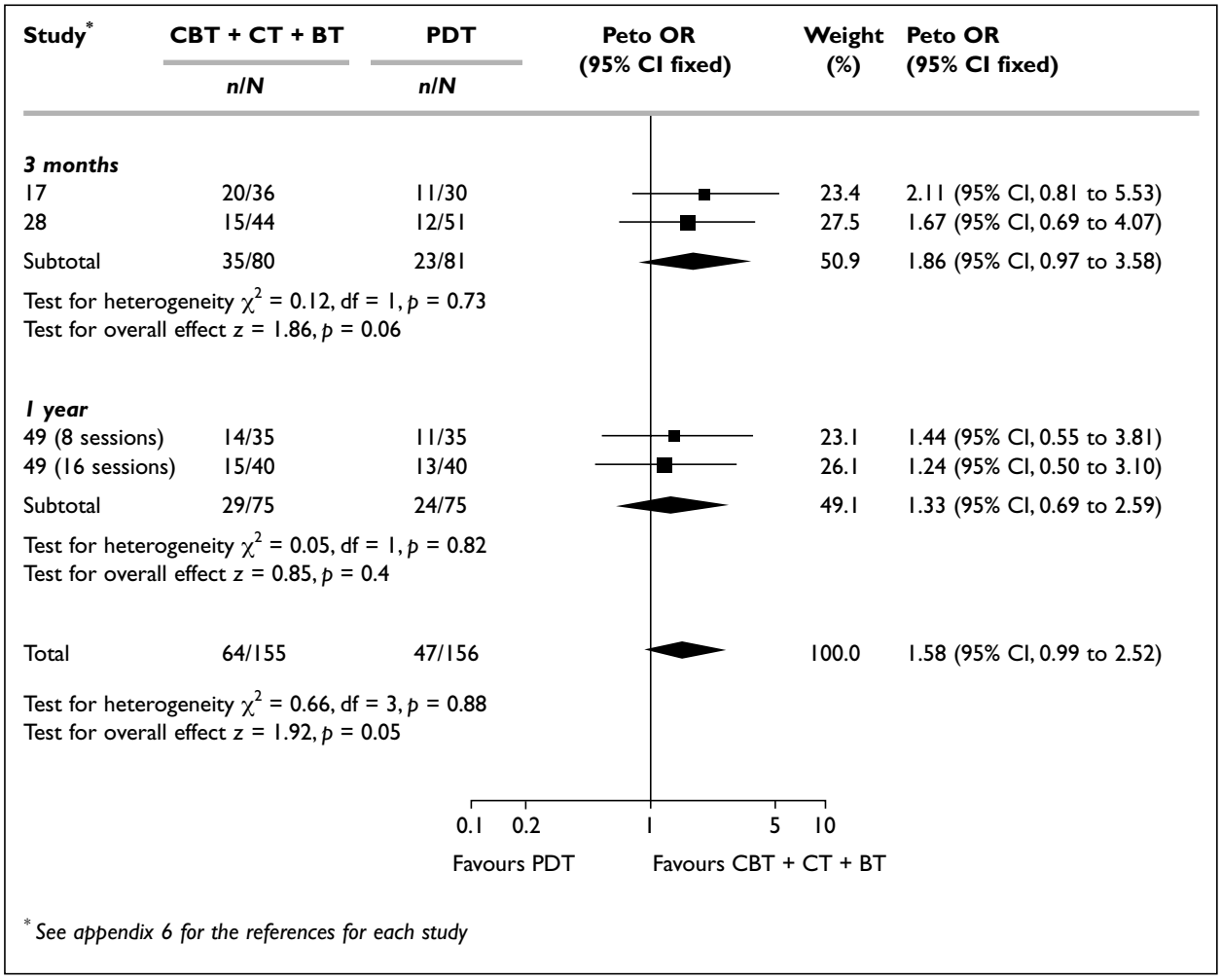


FIGURE 36 *Mean change from baseline at post-treatment for CBT + CT + BT versus PDT*



* *See appendix 6 for the references for each study*

FIGURE 37 *Number of patients who were non-symptomatic at follow-up for CBT + CT + BT versus PDT*

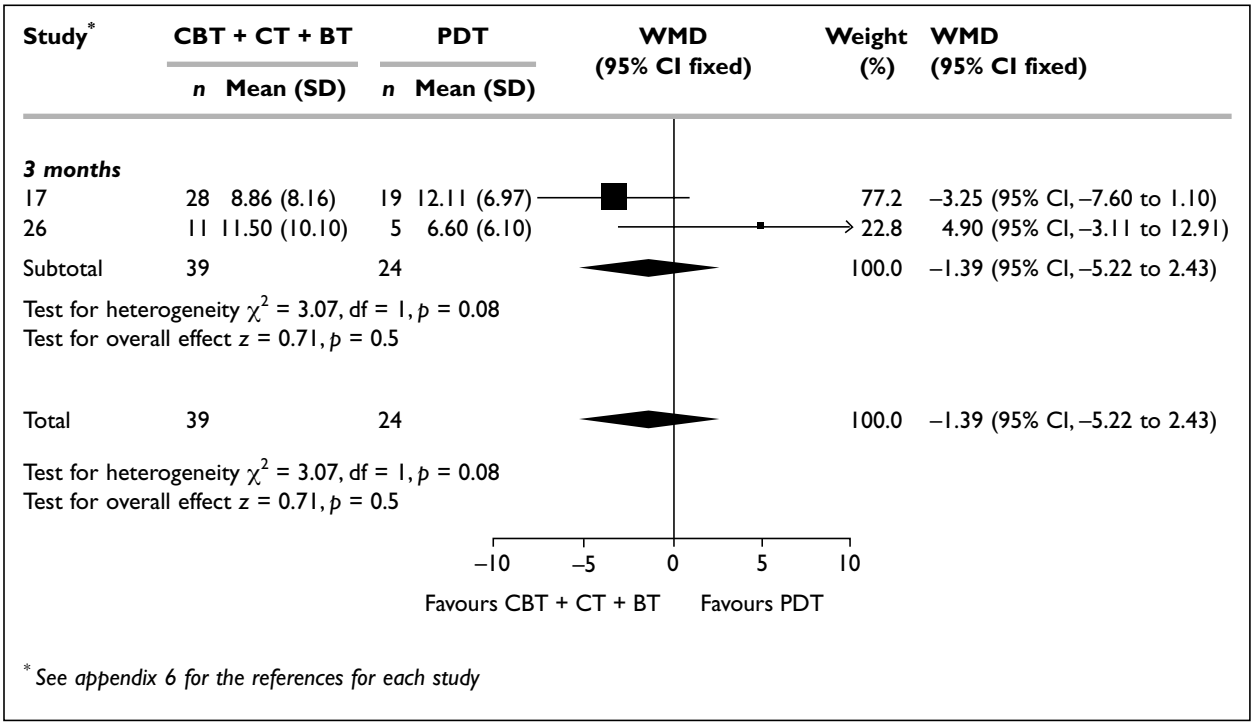


FIGURE 38 *Mean differences at follow-up for CBT + CT + BT versus PDT*

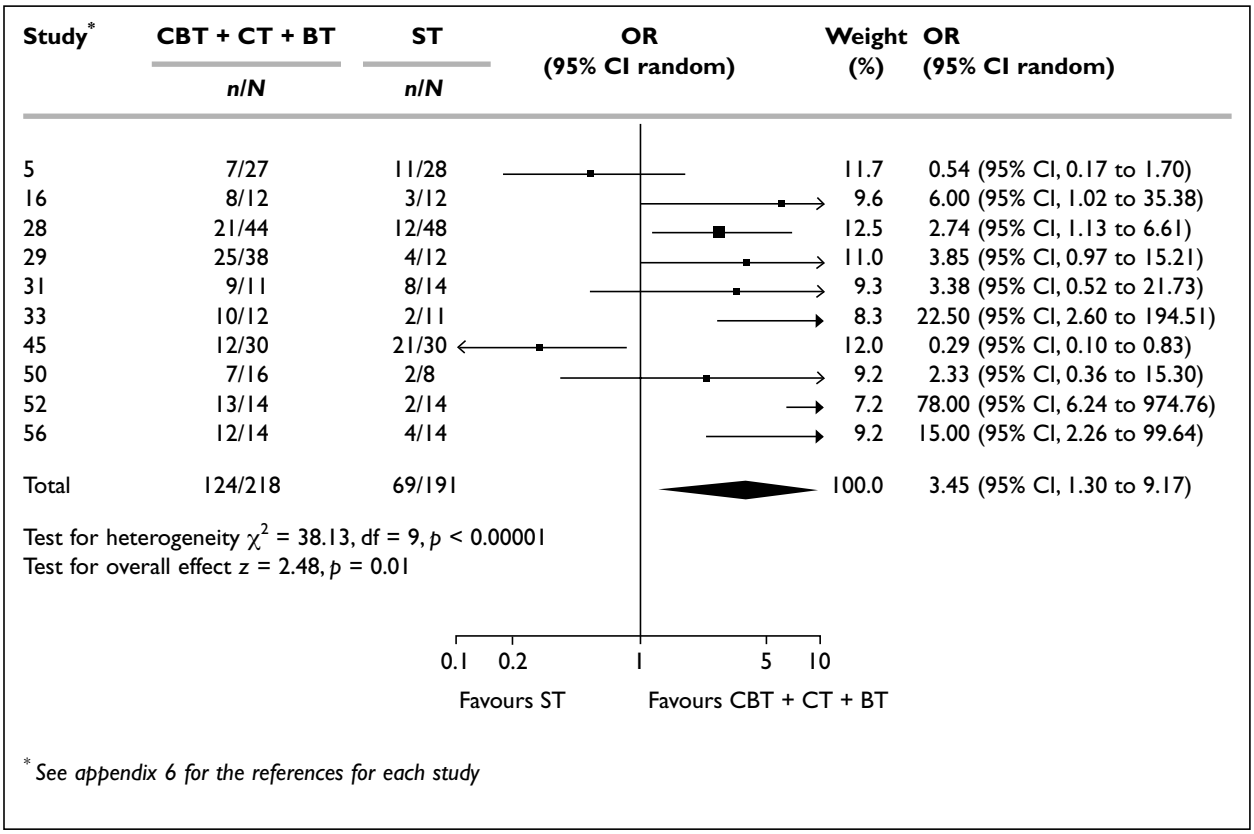


FIGURE 39 *Recovery at post-treatment for CBT + CT + BT versus ST*

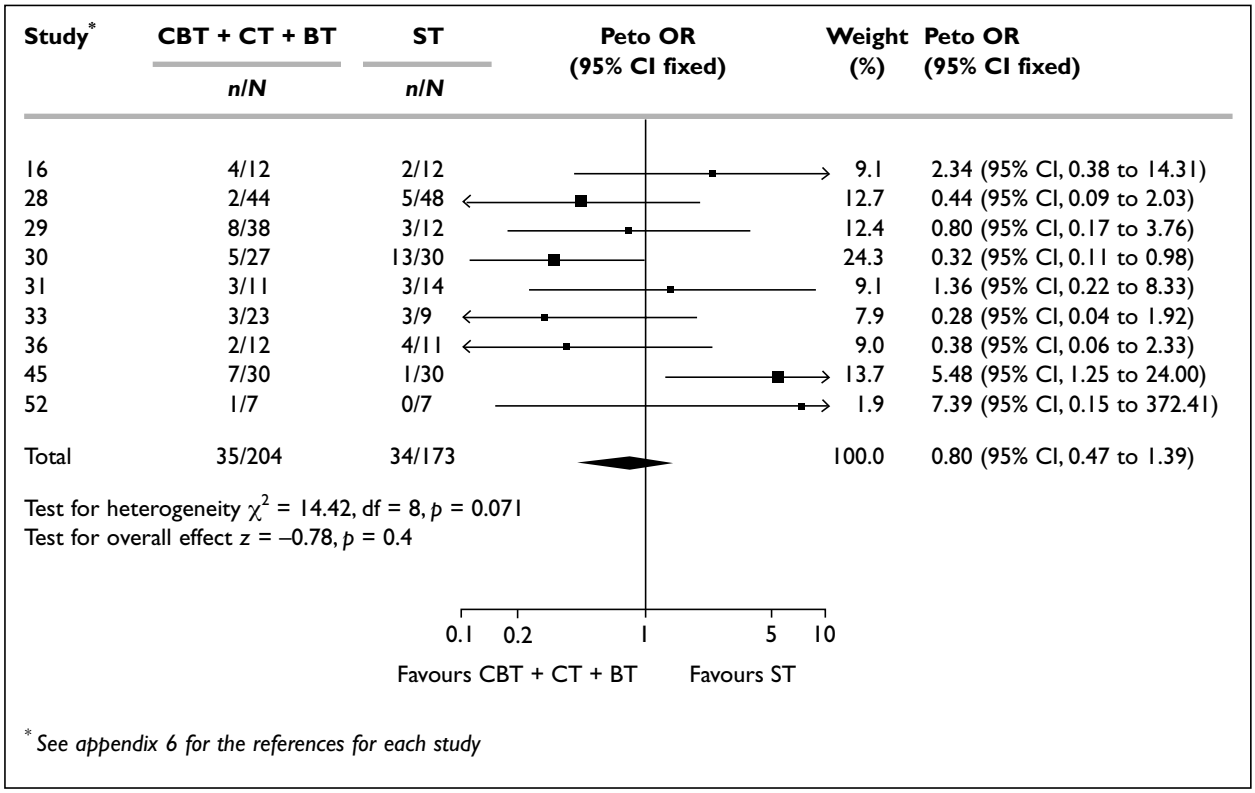


FIGURE 40 *Reported dropouts at post-treatment for CBT + CT + BT versus ST*

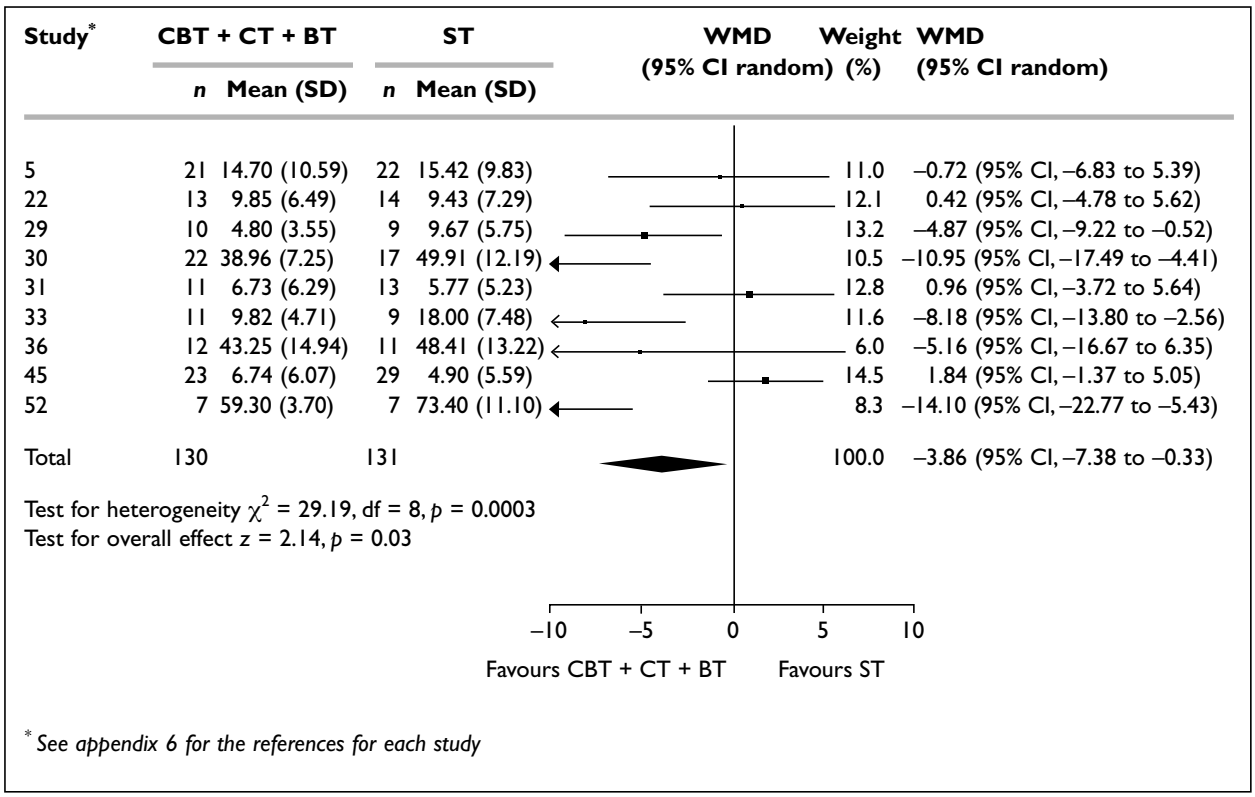


FIGURE 41 *Mean differences at post-treatment for CBT + CT + BT versus ST*

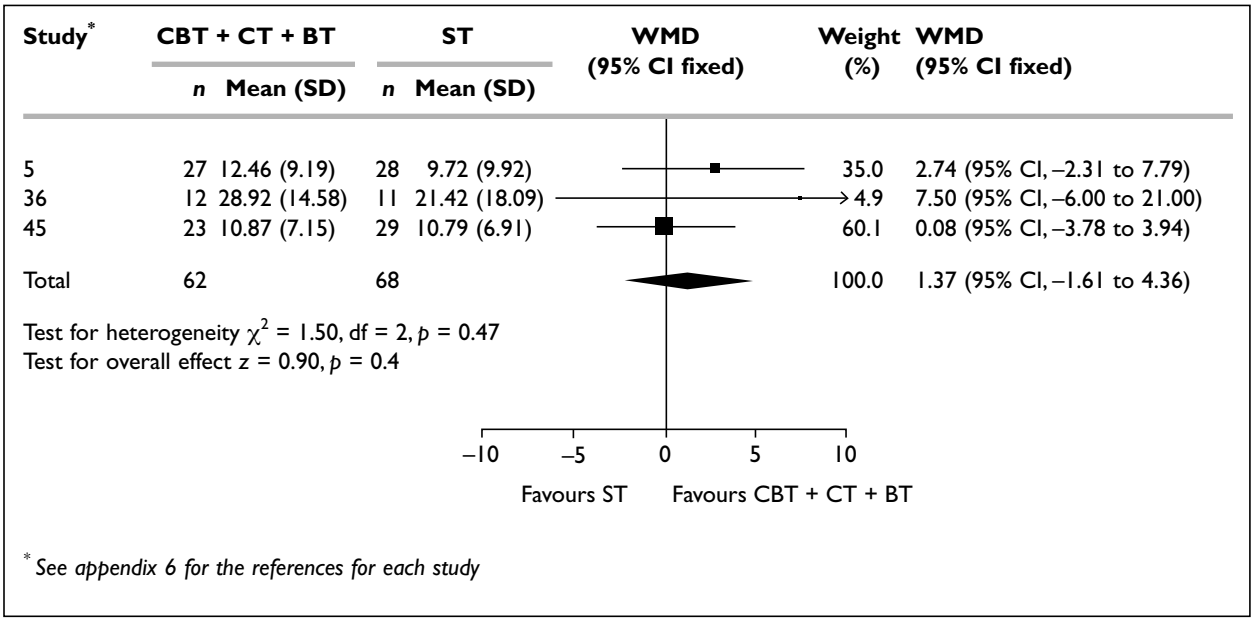


FIGURE 42 *Mean change from baseline at post-treatment for CBT + CT + BT versus ST*

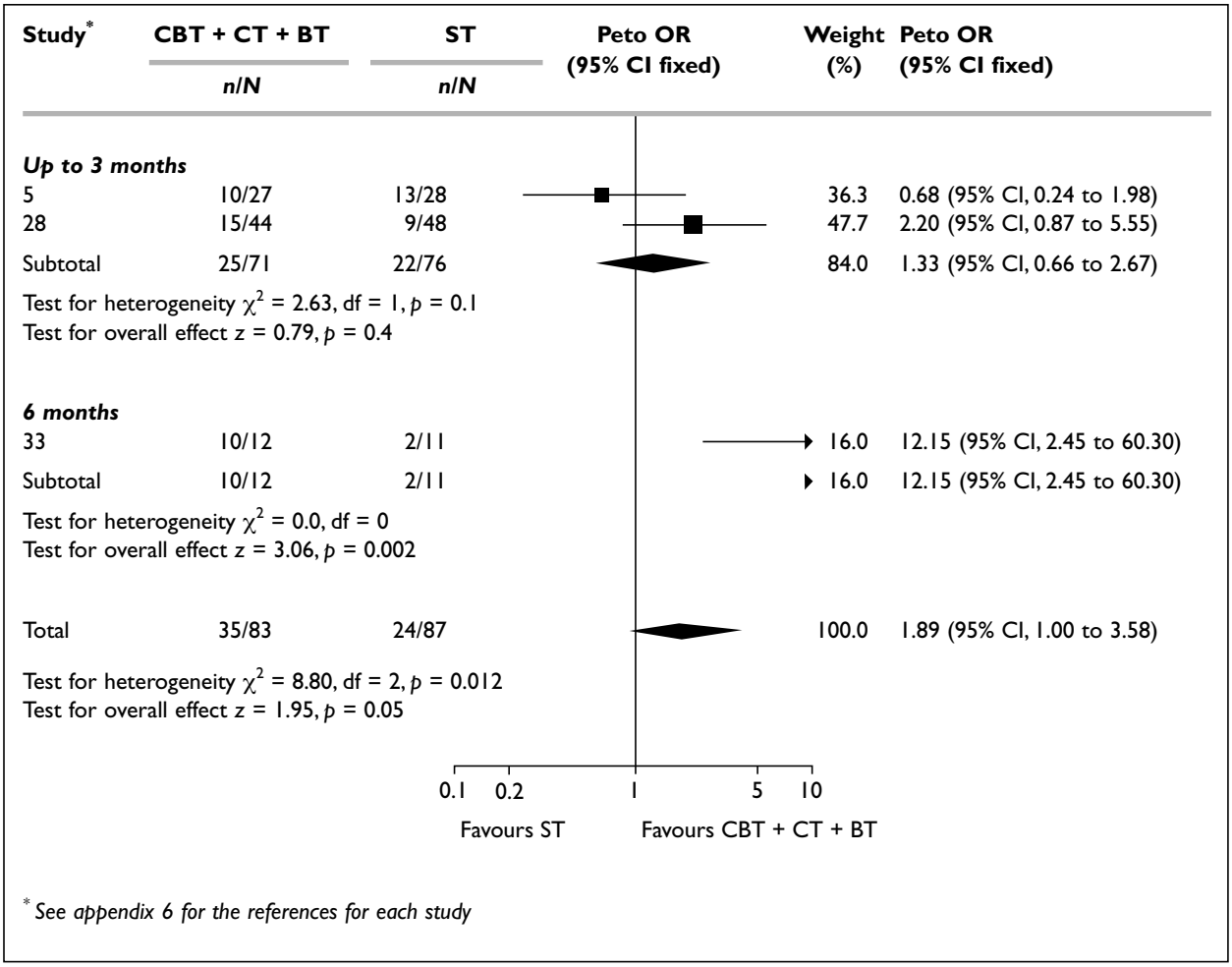


FIGURE 43 *Number of patients who were non-symptomatic at follow-up for CBT + CT + BT versus ST*

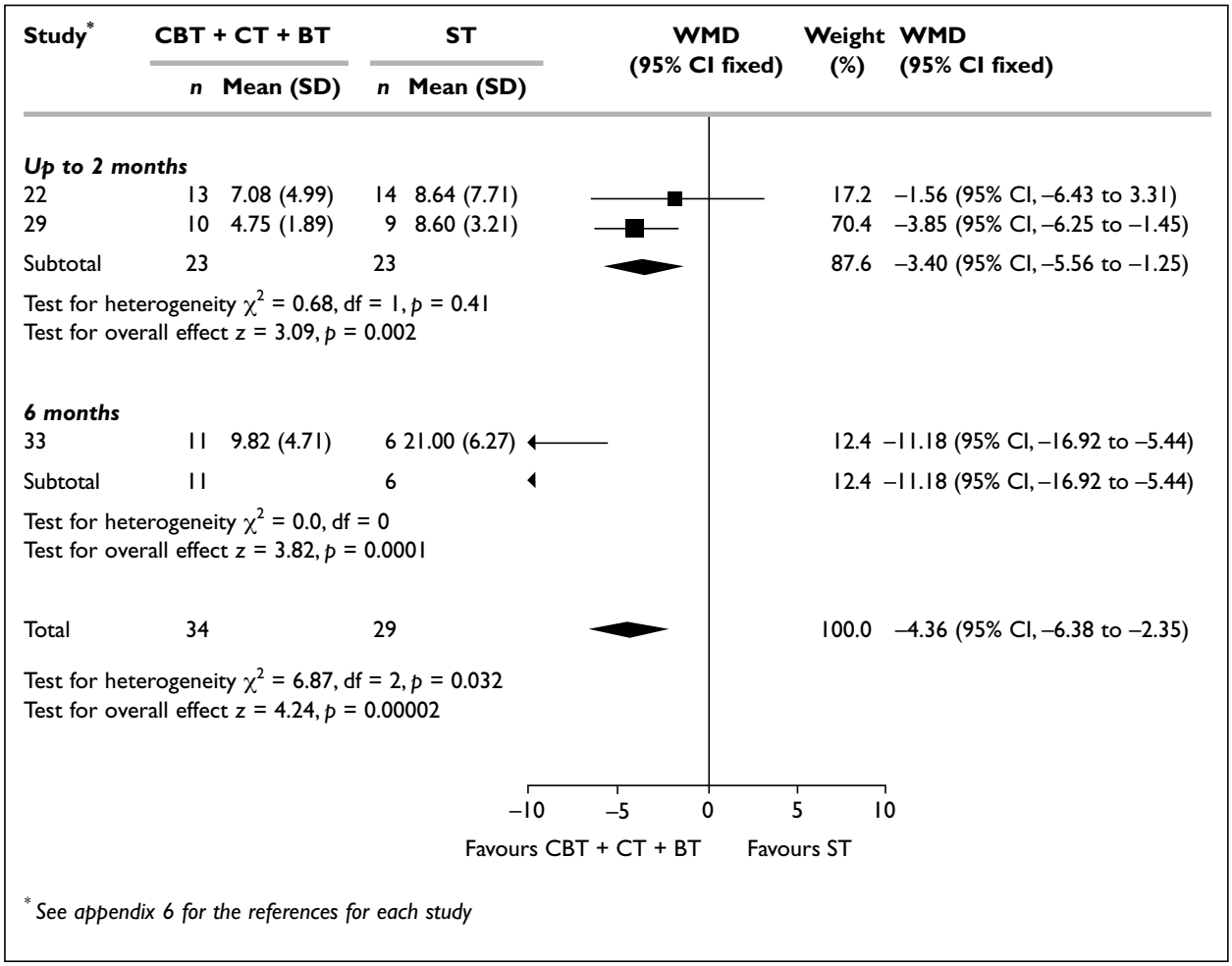


FIGURE 44 *Mean differences at follow-up for CBT + CT + BT versus ST*

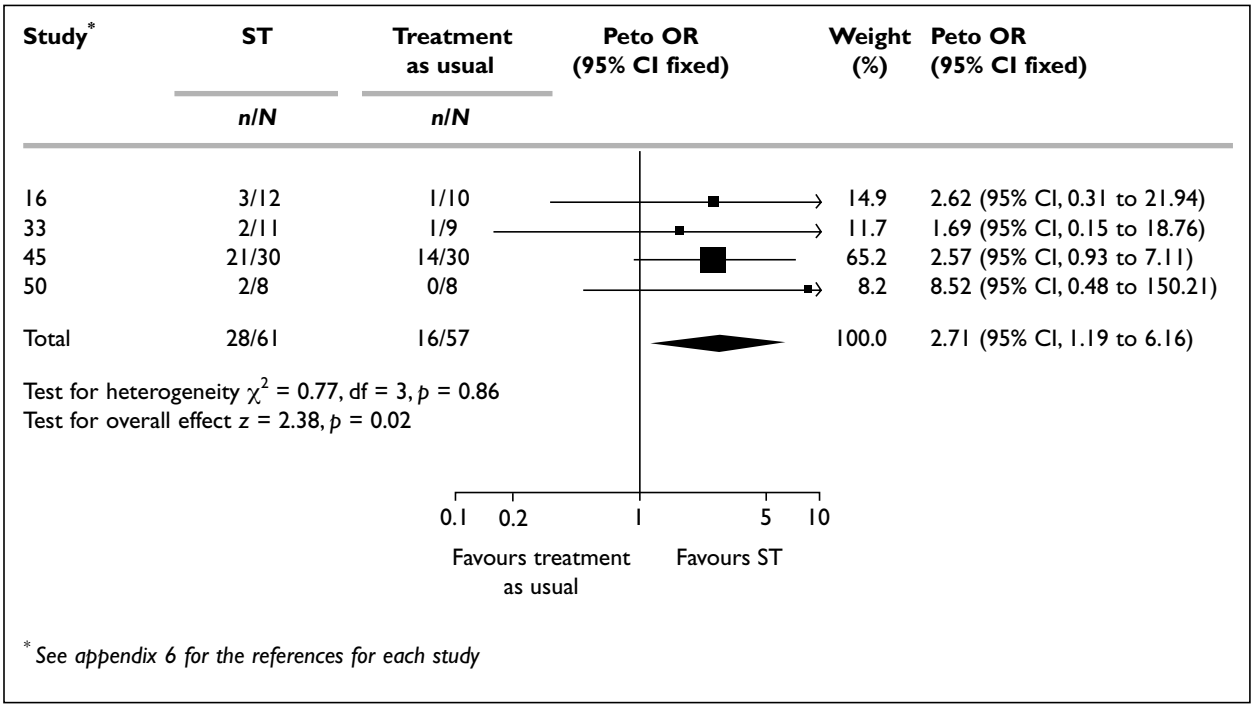


FIGURE 45 *Recovery at post-treatment for ST versus treatment as usual/a waiting-list control*

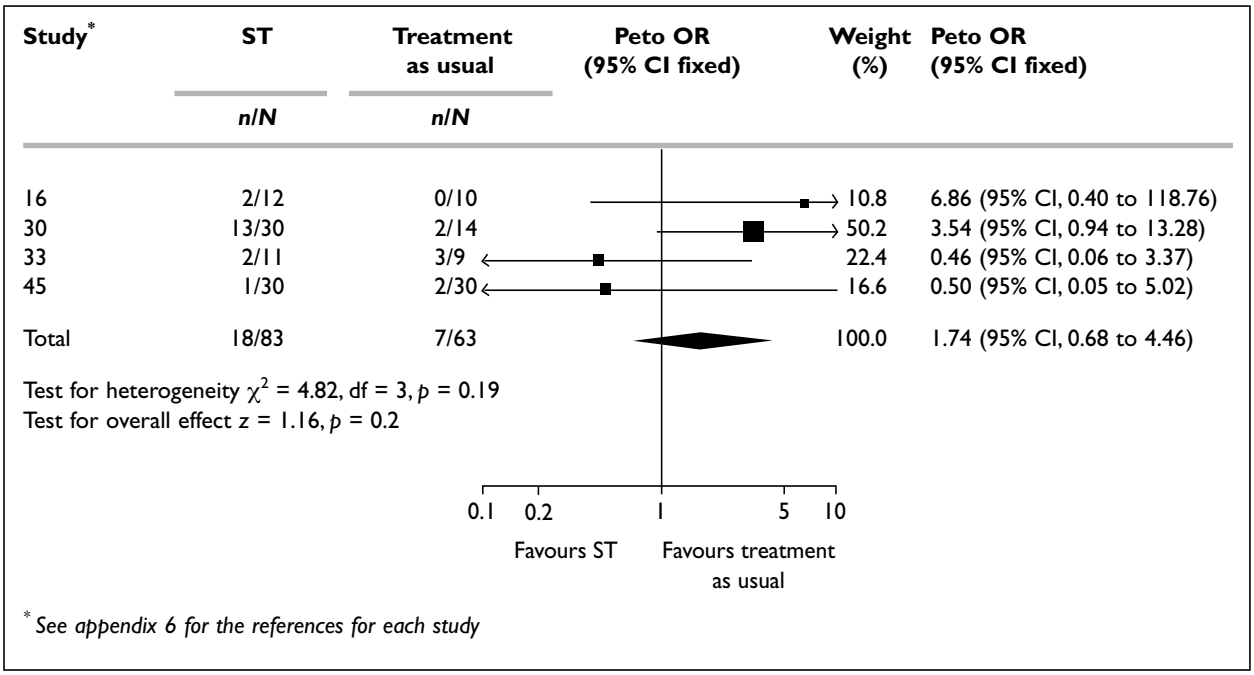


FIGURE 46 *Reported dropouts at post-treatment for ST versus treatment as usual/a waiting-list control*

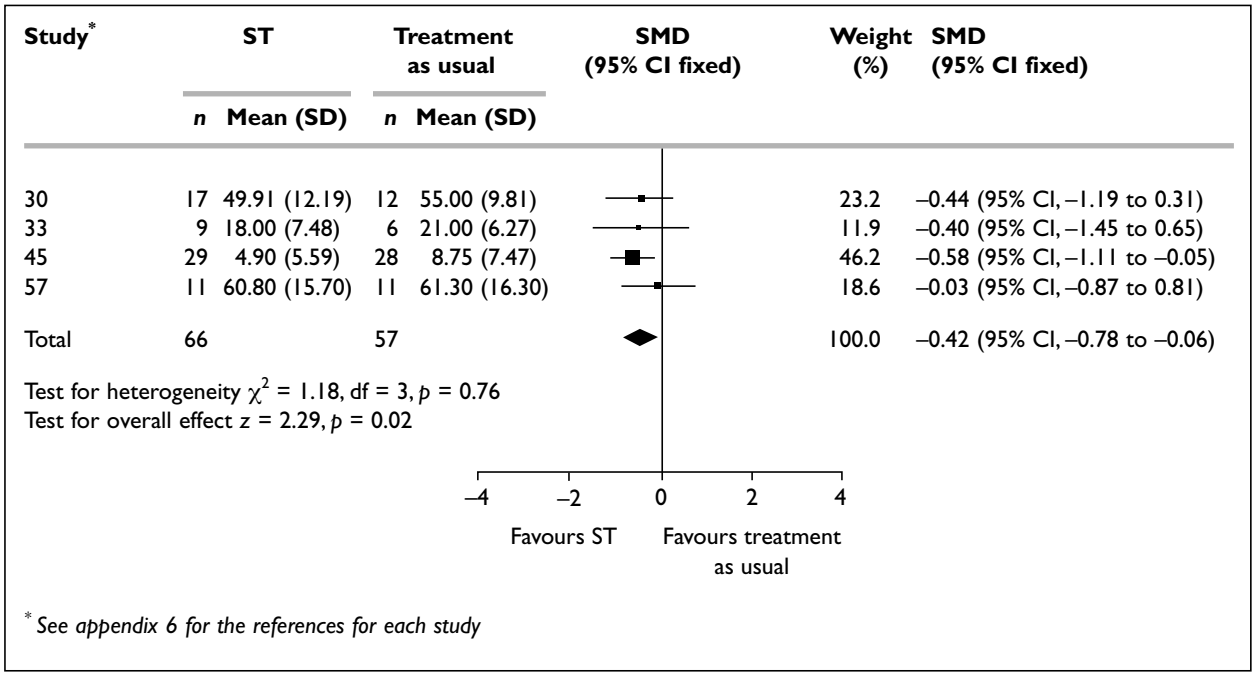


FIGURE 47 *Mean differences at post-treatment for ST versus treatment as usual/a waiting-list control*

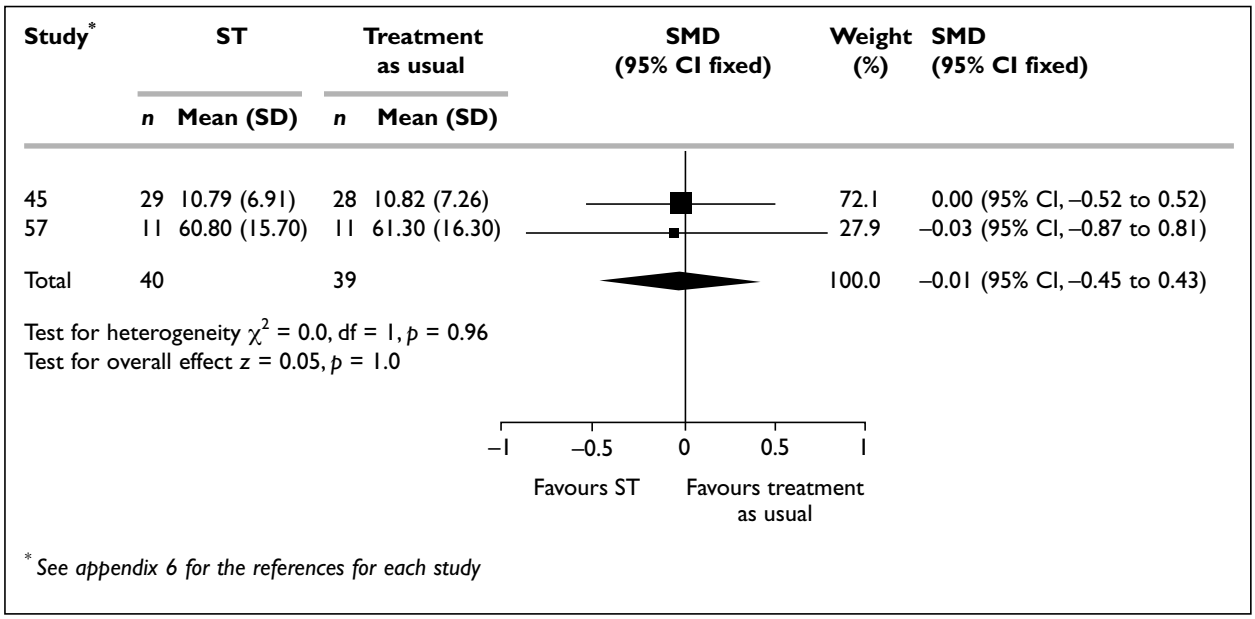


FIGURE 48 *Mean change from baseline at post-treatment for ST versus treatment as usual/a waiting-list control*

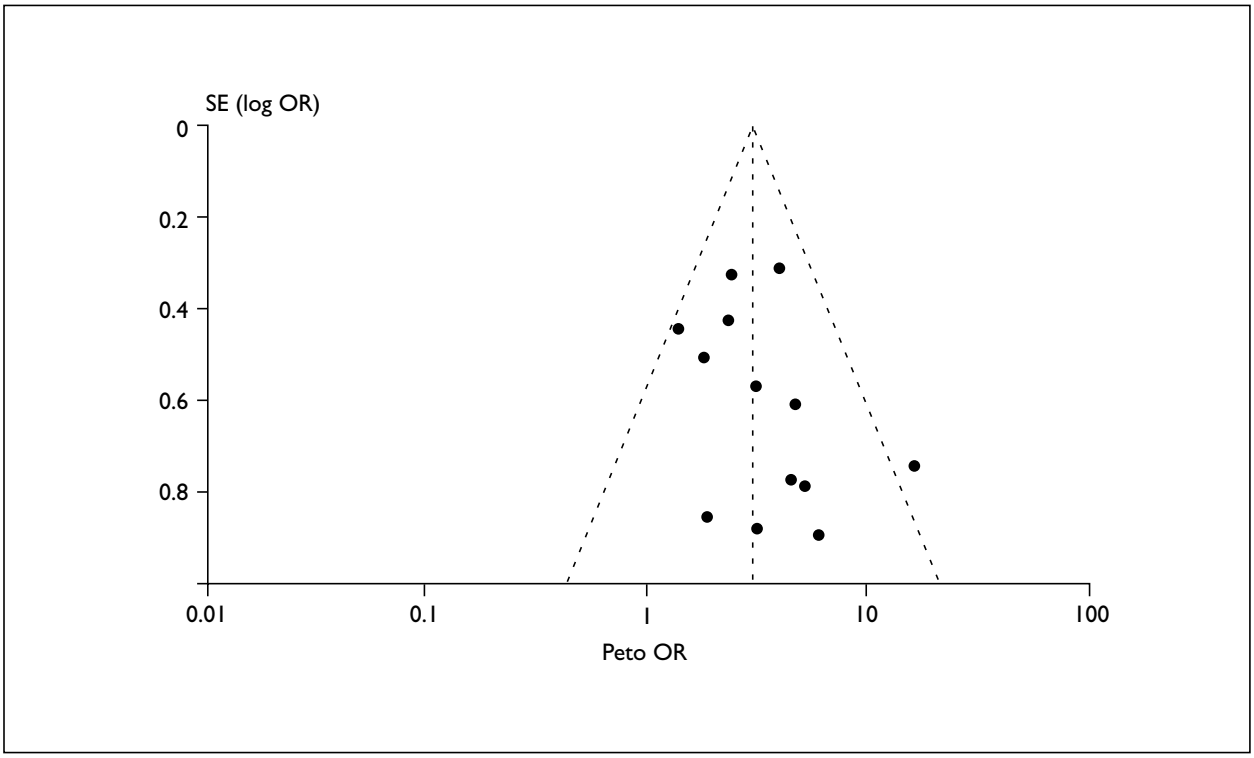


FIGURE 49 *Funnel plot of recovery at post-treatment for all variants of psychotherapy versus treatment as usual/a waiting-list control*

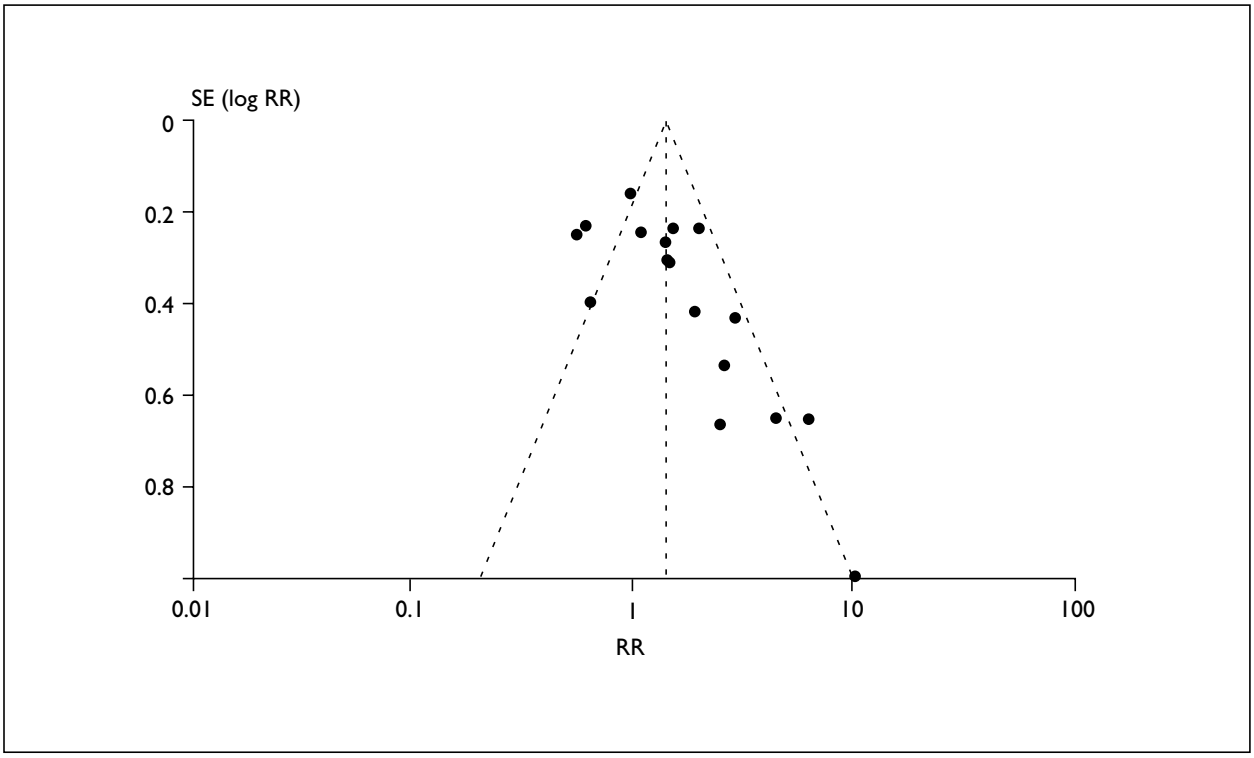


FIGURE 50 *Funnel plot of recovery at post-treatment for CBT + CT + BT versus IPT*

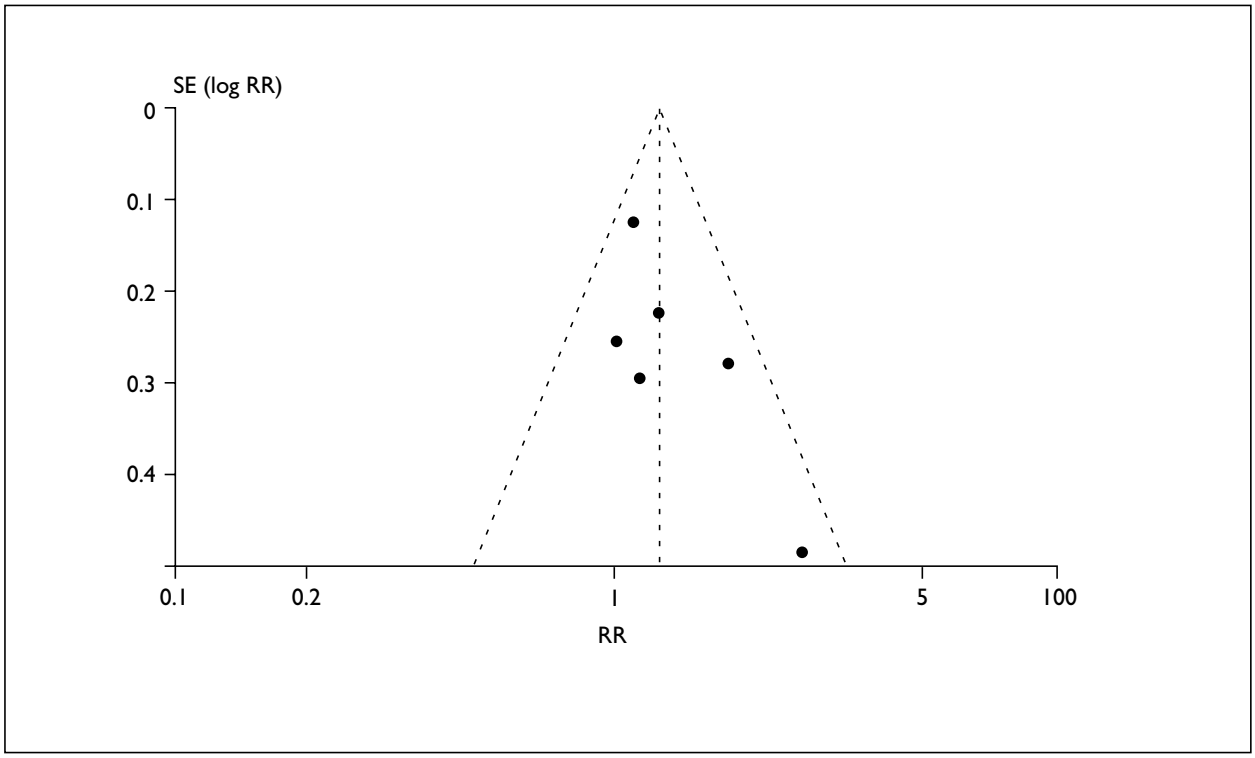


FIGURE 50 *Funnel plot of recovery at post-treatment for individual versus group therapy (all variants)*

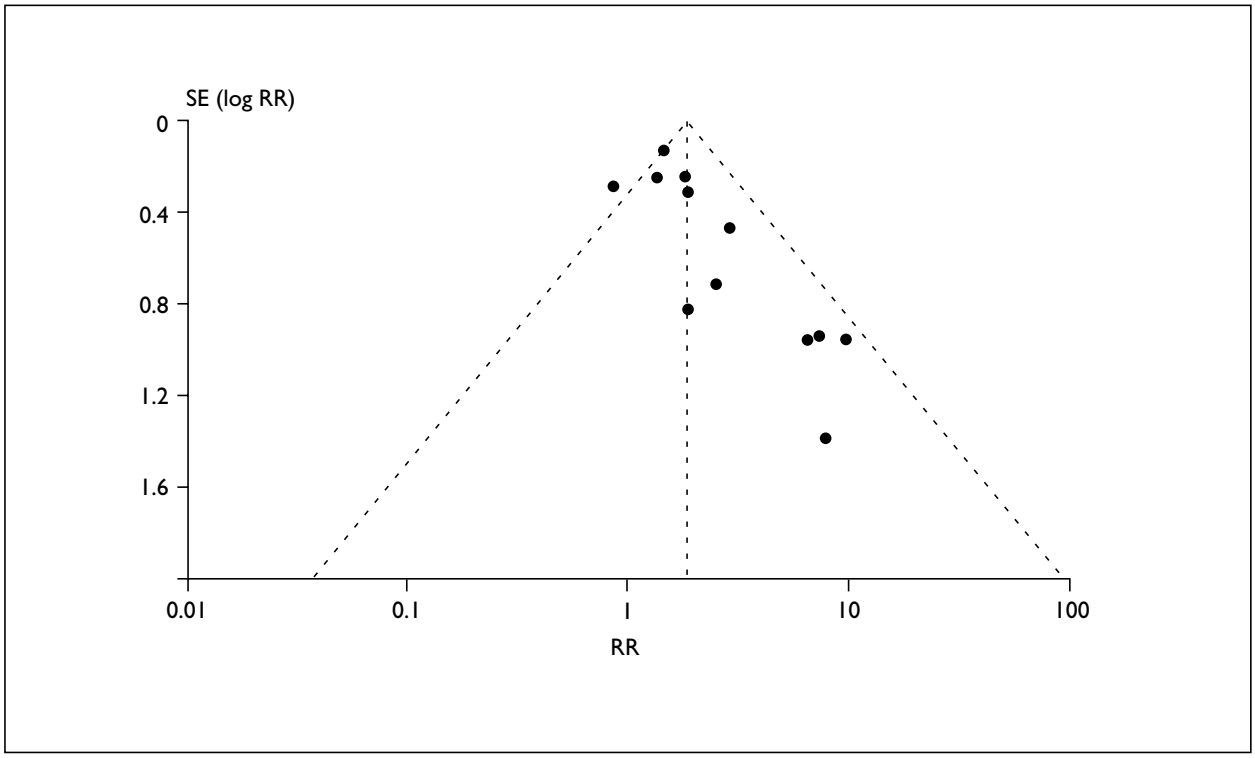


FIGURE 52 *Funnel plot of recovery at post-treatment for CBT + CT + BT versus treatment as usual/a waiting-list control*

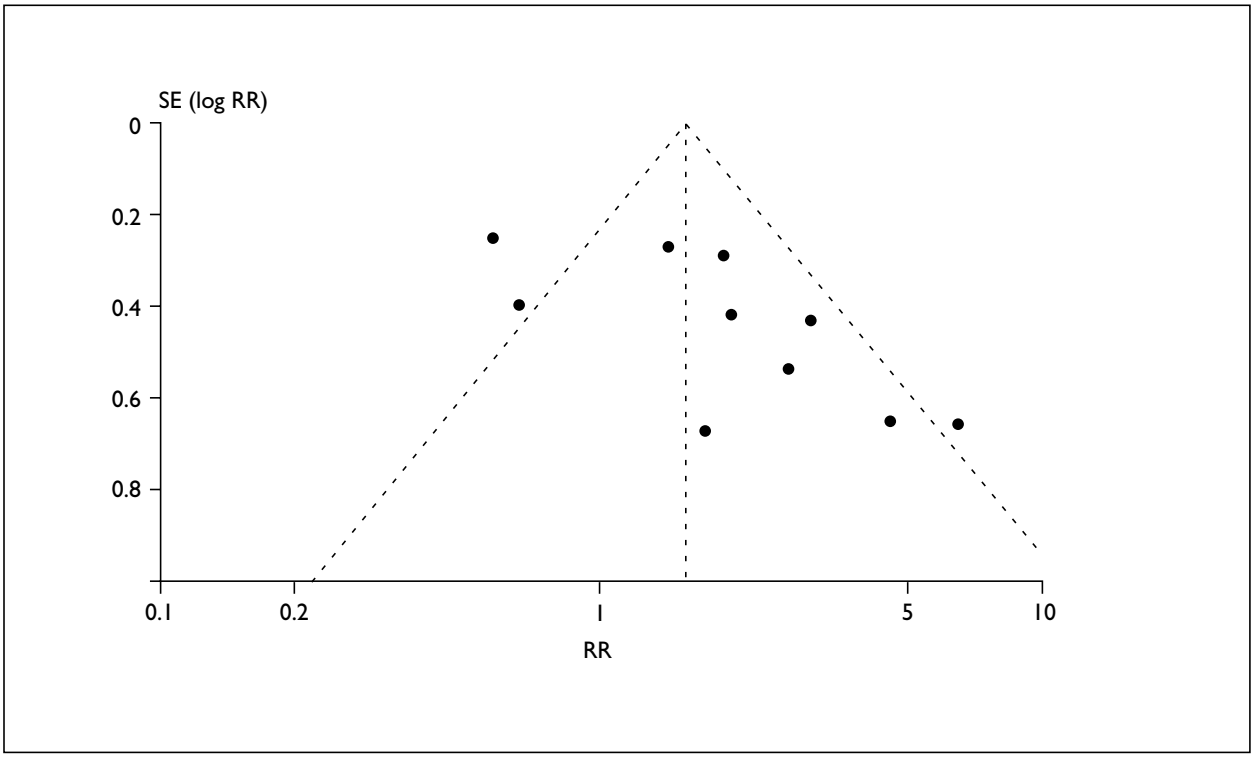


FIGURE 53 *Funnel plot of recovery at post-treatment for CBT + CT + BT versus ST*

Chapter 4
Discussion

The prevalence of depression is high, and the disorder is associated with considerable disability and suffering amongst patients and their families. The first-line treatment for depression is antidepressant medication, although non-pharmacological interventions are becoming increasingly available, and many patients with depression now express a preference for non-pharmacological interventions. In recent years, the provision of psychological therapies in primary care settings has increased, and some concern has been expressed about the lack of evidence supporting their widespread use. With increasing emphasis on evidence-based practice in the NHS, the requirement to demonstrate the effectiveness and cost-effectiveness of psychological treatments is of considerable importance. The primary aim of this systematic review and meta-analysis was to conduct a comprehensive and rigorous assessment of the best available evidence to establish whether brief psychological treatments might offer an effective intervention for patients with depression in primary care. Comparisons between different models of psychotherapy were also undertaken, and the differences in efficacy between individual and group delivery of interventions were investigated.

Internal validity of the trials

The evaluation of psychotherapy poses certain methodological challenges that are not easily addressed within the context of an RCT. In the design of an RCT, the primary aim is to eliminate the potential for bias. This is best achieved by randomisation and blinding patients and investigators to treatment allocation. Whilst allocation concealment is perfectly feasible in RCTs of psychological treatments, blinding of therapists can clearly not be achieved by the inclusion of an 'inactive placebo psychotherapy arm'. In addition, despite the increased use of manuals providing guidance for the use of psychotherapeutic techniques, individual therapist characteristics cannot be controlled in the operationalisation of psychotherapy models, nor can the nature of the therapeutic encounter be predetermined or measured with absolute precision. Furthermore, as in many drug trials, patients are often very well informed about the types of interventions that are available and what they entail. The nature of psychotherapy requires active participation by patients, and it is possible that at least some of them would be able to identify prominent theoretical constructs during therapy, thereby introducing potential bias on the part of the patient. Finally, research in this area has so far failed to determine, with any certainty, single or multiple specific 'active' components of psychotherapy, and, since the mechanism of action in most types of psychotherapy is largely theoretical, there can be significant problems in identifying an appropriate control or comparison condition. Given these caveats, it is still possible to conduct a well-designed controlled trial to examine the efficacy of psychological interventions. However, low scores on internal validity items were recorded for all but a small number of trials in this review. This, coupled with the inadequate reporting of methodology resulted in low overall quality scores. In addition, it should be emphasised that, although 63 studies were eligible for inclusion in the review, 12 studies provided insufficient data for inclusion in the meta-analyses (studies 1, 2, 8, 10, 15, 18, 20, 24, 27, 32, 37 and 61*).

Randomisation procedure

Randomisation and concealment of allocation have been shown to influence outcomes in RCTs. Indeed, allocation concealment has been found to be more important in preventing bias than the generation of allocation sequence. However, although 54 trials in this review (86%) reported their assignment procedure as being randomised (studies 1–6, 8–23, 25, 27–35, 37–40, 42–49, 53–55, 57–63*), only ten of those trials described either the randomisation or allocation concealment procedures (studies 5, 13, 17, 21, 25, 29, 31, 40, 45 and 47*). This paucity of information suggests bias may have been introduced during the allocation procedure in the majority of the trials.

* See appendix 6 for the references for each study.

Type of comparison arm

Just under 50% of trials used treatment as usual or no treatment control groups, of which two-thirds used a waiting-list arm in order to overcome ethical concerns about withholding treatment from one group of subjects (studies 2, 6–9, 16, 21, 22, 24, 30, 33, 34, 38, 40, 43, 48, 50, 51, 53, 58 and 60*). Some researchers consider that waiting-list control groups may have an unintended reverse placebo effect because an implicit suggestion is being made to participants that they should not expect to improve until treatment begins.[61] In this way, waiting-list controls could have more negative outcomes than would have occurred through the passage of time, and the potential for bias in favour of the psychotherapy treatment groups might be considerable. On the other hand, it is possible that being on a waiting list may have a positive effect because the length of time in this group is finite and therapeutic intervention is guaranteed.

Concomitant/adjunctive treatments

The variable use of antidepressants in combination with psychotherapeutic interventions is of particular concern in assessing the evidence for the effectiveness of psychological treatments. Investigators of 24 trials (37%) specifically requested that no patients take antidepressant medication alongside psychotherapy (studies 5, 11–13, 17, 20, 23, 26, 28, 31–36, 38, 39, 53, 56, 59, 62 and 63*), but nearly one-fifth of trials allowed naturalistic prescribing to take place in any of the treatment groups (studies 1, 25, 27, 30, 42, 46, 47, 49, 54 and 60*), and many other authors simply failed to report any policy of additional prescribing (studies 2, 3, 6–9, 14–16, 18, 19, 21, 22, 24, 29, 37, 43, 48, 50–52, 55, 57 and 61*). Researchers evaluating the coping with depression course[48] (studies 6 and 30*) allowed patients to receive concurrent psychotherapy treatment independently of the trials in which they were participating. In view of both this inadequate reporting and the apparent variability in treatment administration, there can be no certainty that reduction or recovery from depression was achieved through psychotherapeutic intervention by trial therapists alone. It should be noted that, in the trial conducted by Katon and co-workers (study 25*), patients were required to take antidepressants as one of the inclusion criteria, since the primary aim of the psychological intervention was to improve adherence to medication, with recovery from depression used as a secondary outcome.

ITT and adherence to treatment

Only a small number of studies (17%) used an ITT analysis, in which missing data were substituted with the last available observation (studies 23, 25, 36, 40, 43–45, 47, 48, 51 and 59*). Treatment dropouts and follow-up failures, both known and unknown, are likely to be systematically different in many ways and could influence treatment outcomes. Thus, the principle upon which the validity of the RCT design rests (equivalence in every respect other than the allocated intervention) is no longer maintained. More than half the trials included in the review excluded randomised patients who did not commence treatment or who later dropped out, and post-treatment comparability between groups could not be assumed in these trials (studies 1, 2, 6–9, 14–17, 20, 21, 26–30, 33–35, 37, 39, 41, 42, 46, 49, 54–58 and 60–63*).

Sample size

A further concern in assessing the strength of the evidence is that the majority of trials contained very small sample sizes, with a median arm size for all trials of just 13 patients. Multiple small studies increased the likelihood of both type 1 and type 2 errors. Only seven trials (studies 1, 5, 13, 20, 25, 45 and 50*) made reference to power calculations, and just two trials (studies 13 and 25*) had sufficient statistical power to detect a meaningful treatment difference between two psychotherapy models (it should be noted that trials that compared two 'active' psychotherapy models, would probably require larger numbers of patients in each arm than trials comparing one model with treatment as usual).

Outcome measures

The majority of trials (84%) used the BDI to measure outcomes in terms of recovery from depressive symptoms (studies 1–9, 11, 13–18, 19, 20–24, 26–35, 37–43, 44, 46–50, 53–55, 58–60, 62 and 63*). The BDI is a widely used self-rated instrument and avoids the potential pitfalls of subjective clinician-rated measures. However, it is limited to the measurement of symptomatic clinical outcomes. Patients with depression are disabled in many spheres of activity, and broader measurements of levels of functioning, such as quality-of-life scales, might be more intuitively meaningful and sensitive to changes in patients receiving psychological treatments. Quality-of-life scales measure several dimensions, including social and occupational functioning, role per-

* See appendix 6 for the references for each study.

formance and physical symptoms and, furthermore, put clinical changes into perspective. It is disappointing to note, however, that quality-of-life outcome measures rarely featured in the trials of this review. Patient satisfaction would have been another useful treatment outcome measure, but just six trials included such scales in their battery of tests (studies 3, 5, 6, 14, 20 and 42*).

Researcher bias

It was evident that concomitant with the development of CT in the early 1970s, CT and BT models have been subjected to regular evaluation over the last 25 years. Indeed, more than 90% of trials in the review included at least one CT, BT or CBT arm. In contrast, very few trials examined the effectiveness of PDT or client-centred therapy. Where client-centred or traditional insight-oriented therapy were included (studies 5, 20, 29, 30, 36, 50, 56 and 57*), they tended to be regarded and utilised by investigators as an attention-placebo control, and detailed descriptions of theoretical foundations and techniques were usually omitted. This raises the possibility of researcher bias. In the majority of trials, the authors had developed, or been closely involved in the development of the therapy under evaluation. It is unlikely that participating therapists undertook all psychological treatments with equal levels of training, knowledge, skills and commitment. This might be one reason why PDT appeared to be less efficacious than CBT models, although it is noteworthy that ST appeared to perform reasonably well under similar conditions.

Application of therapeutic technique

Although many authors reported using manuals to standardise individual psychotherapy interventions, only one-sixth of trials monitored the psychological intervention through weekly supervision discussions with the therapists (studies 6, 16, 33, 34, 37, 38, 41, 55, 59 and 61*). Nearly 40% of the trials included in the review failed to monitor adherence to the psychotherapy interventions under evaluation (studies 2, 4, 7, 10–12, 14, 18, 19, 21, 24, 27, 35, 40, 42, 45, 47, 48, 51–53, 56, 60 and 63*). Therefore, it cannot be assumed that the therapists in those trials consistently applied the models as directed, and observed outcomes cannot be attributed with complete certainty to the effects of the models themselves.

Interpretation of the findings

Psychotherapy (all variants) versus treatment as usual

All 13 trials contributing to the post-treatment recovery analysis were reportedly randomised, although the method used was often not described. All but one (involving IPT) compared some variant of CBT with treatment as usual or a waiting-list group. Seven of the trials were conducted in the USA (studies 16, 25, 33–35, 44 and 48*), four in the UK (studies 40, 45 and 46 and 54*), one in Canada (study 50*) and one in Australia (study 58*). For the recovery outcome, no statistical heterogeneity was apparent. Pooling these data produced a highly significant and extremely robust result, indicating that patients in receipt of psychotherapy were significantly more likely than those on waiting lists or receiving treatment as usual to improve to a degree where they were no longer regarded as being clinically depressed. Subgroup analyses suggested that this effect might be moderated by differences in baseline severity (trials involving more severely depressed patients demonstrated a less marked difference in favour of psychotherapy) and possibly by differences in the number of sessions offered (trials involving more sessions demonstrated a greater difference in favour of psychotherapy).

All but two (studies 7 and 51*) of the 22 trials contributing to the analysis of the post-treatment mean differences and all six contributing to the analysis of the mean change post-treatment were reportedly randomised, although in the majority of cases the method used was not described. In these trials reporting continuous data, all but two (study 44 involving IPT and study 57 involving ST*) compared some variant of CBT with treatment as usual or a waiting-list group. Of those trials reporting mean differences, 16 were conducted in the USA (studies 3, 6, 7, 9, 14, 25, 33–35, 38, 43, 44, 48, 51, 53 and 57*), three in the UK (studies 40, 45 and 46*), one in Canada (study 30*) and two in Australia (studies 58 and 60*). Three of the trials reporting data on the mean changes were conducted in the USA (studies 21, 44 and 57*), two in the UK (studies 45 and 46*) and one in Australia (study 58). Statistical heterogeneity was highly significant in the mean differences data, although none was apparent in the mean change data. The results obtained by pooling dichotomous outcomes were broadly supported by the findings from the continuous data. These pooled

* See appendix 6 for the references for each study.

continuous data suggested that those treated with psychotherapy exhibited significantly fewer symptoms post-treatment than those on waiting lists or receiving treatment as usual, and that symptom reduction from baseline was significantly greater in those receiving psychotherapy. Although the data from dichotomous and continuous analyses broadly supported one another, when the mean differences data were re-analysed using only ten trials that reported both types of data (studies 25, 33–35, 40, 44–46, 48 and 58*), pooling resulted in an increase in the observed statistical heterogeneity. Subgroup analyses failed to remove the observed statistical heterogeneity, but suggested that the overall effect might, again, be moderated by differences in baseline severity (trials involving more severely depressed patients demonstrated a less marked difference in favour of psychotherapy) and differences in the number of sessions offered (although there was no obvious trend and this analysis was more difficult to interpret). In all subgroups, the pooled effect size remained highly statistically significant.

Only four trials contributed dichotomous data on the number of patients that were non-symptomatic at follow-up (at 3 months only). All compared some variant of CBT with treatment as usual or a waiting-list group. Three of the trials were conducted in the USA (studies 25, 35 and 48*) and one in the UK (study 54). Statistical heterogeneity was significant in this outcome, primarily due to one small trial with an extreme effect size. This trial was different from the others in that it had involved people with unspecified levels of depression. Removal of this trial reduced the heterogeneity to a level that was non-significant. The differences observed at post-treatment were not demonstrated at 3-month follow-up and pooling of dichotomous data suggested no significant difference between the two groups.

Of the 11 trials contributing to the pooled mean differences at follow-up (at 1, 3, 6 to 9 months and 1 and 2 years), all but two (study 44 involving IPT and study 57 involving ST*) compared some variant of CBT with treatment as usual or a waiting-list group. All were conducted in the USA, except one that was carried out in the UK (study 46*). Three of the four trials contributing to the pooled mean change at follow-up data (3, 6 to 9 months and 1 year) were conducted in the USA (studies 25, 44 and 57*) and one in the UK (study 46*). Statistical heterogeneity in the mean differences data was highly significant only at 1-month follow-up and this was apparently due to the effects of one trial. No statistical heterogeneity was observed at any other timepoint and none was apparent at any timepoint in the mean change data. Pooled mean differences at each timepoint suggested that those treated with psychotherapy exhibited significantly fewer symptoms than those on a waiting list or those receiving treatment as usual at 1 and 3 months and between 6 and 9 months. Data could not be pooled for 1- and 2-year follow-up. Symptom reduction from baseline was also significantly greater in those receiving psychotherapy at 3 months and between 6 and 9 months.

Although the number of dropouts appeared to be slightly greater in the psychotherapy group, there appeared to be no significant differences in dropouts between those receiving psychotherapy and those on waiting lists or receiving treatment as usual. Dropouts appeared to be greater in the psychotherapy group only in those trials involving patients with major depression, although the difference did not reach significance. Summarising the available dropout data suggested a dropout rate of 20% from the psychotherapy group and 15% from treatment as usual. Since dropouts were more likely to be unreported and we have assumed all missing data to be treatment failures, it is possible that we have under-estimated the benefits of psychotherapy.

Although significant differences were still observed in all categories, the effect in favour of psychotherapy in the recovery and mean differences data was more extreme in those trials recruiting self-selected patients and volunteers than in those involving clinic-attenders and outpatients. However, statistical heterogeneity was still apparent in both categories in the mean differences data. No differences between trials according to the source of recruitment were demonstrated in the dropout data. The overall quality score of the trial had little bearing on outcome for recovery or dropouts, but for pooled mean differences, for which more data were available, higher-quality trials appeared to result in a lower effect size than lower-quality trials. A funnel plot drawn from the data of this comparison suggested evidence of bias by demonstrating the absence of small 'negative' trials. Thus, once more, the importance of interpreting the findings with caution cannot be over-emphasised.

* See appendix 6 for the references for each study.

This comparison combined all available economic evaluations but none was sufficiently large to allow robust testing of efficiency differences between psychotherapy and treatment as usual. However, each suggested that psychotherapy might have been superior with larger samples. Two of the four cost-effectiveness analyses were generally well designed and well conducted, providing tentative support for the hypothesis that psychotherapy is more efficient than usual care.

CBT versus IPT/PDT/ST

Of the 16 trials (providing 17 data sets) contributing to the post-treatment recovery analysis, 13 (studies 4, 5, 11, 13, 16, 17, 28, 29, 31, 33, 45, 49 (eight-session arm), 49 (16-session arm) and 50[*]) were reportedly randomised (studies 45 and 49 using allocation concealment) and three were CCTs (studies 26, 52 and 56[*]). All trials used formal CT, BT or CBT models. ST and PDT were most often used as the comparison group with IPT being used in only two trials (studies 13, 49 (eight-session arm) and 49 (16-session arm)[*]). Twelve of the trials were conducted in the USA (studies 4, 5, 11, 13, 16, 17, 26, 29, 31, 33, 52 and 56[*]), two in the UK (studies 45 and 49[*]) and two in Canada (study 28 and 50[*]). Statistical heterogeneity was highly significant. Four of the trials (studies 11, 33, 52 and 56[*]), contributing 20% of the weight between them, reported extremely high ORs (> 15) and a single relatively large trial suggested a significant finding in the opposite direction (study 45[*]). The result was highly significant indicating that patients receiving variants of CBT were significantly more likely than those receiving any of the other three models (IPT, PDT or ST) to improve to a degree where they were no longer regarded as being clinically depressed. Subgroup analyses suggested that this effect was moderated by differences in baseline severity. A highly significant difference between groups was observed where severity was unspecified, but no difference was apparent in trials involving more severely depressed patients. Statistical heterogeneity was still apparent in the group of trials involving more severely depressed patients. The overall effect also appeared to be moderated by differences in the number of sessions offered (trials involving fewer sessions demonstrated a greater difference in favour of CBT, an observation in the opposite direction from that in comparison 1) although, once more, statistical heterogeneity was still apparent.

All but three of the 13 trials (studies 26, 36 and 52[*]) contributing to the mean differences post-treatment analysis and all but one (study 36[*]) of the six contributing to the mean change post-treatment analysis were reportedly randomised. All trials used formal CT, BT or CBT models and ST was most often used as the comparison group. Of the trials reporting mean differences, 11 were conducted in the USA (studies 4, 5, 13, 17, 22, 26, 29, 31, 33, 36 and 52[*]), one in the UK (study 45[*]) and one in Canada (study 30[*]). Three of the trials reporting mean change were conducted in the USA (studies 5, 17 and 36[*]) and two in the UK (studies 45 and 49[*]). Statistical heterogeneity was highly significant in the mean differences data. A single trial (study 26[*]) reported a considerable difference between groups in the opposite direction from the majority of the other trials. In addition, although reportedly measuring outcomes on the same scale, some trials (studies 30, 36 and 52[*]) reported unexpectedly high group means. Although no statistical heterogeneity was apparent in the mean change data, one of these trials (study 36[*]) also appeared in this analysis, and, again, reported considerable mean changes in both groups, quite different from the other trials in the group. Although there appeared to be a trend in favour of variants of CBT in both sets of data, the results obtained by pooling dichotomous outcomes were not strongly supported by the findings from continuous data. There were no significant differences between groups in terms of post-treatment symptomatology or symptom reduction from baseline. When the mean differences data were re-analysed using only the ten trials that reported both dichotomous and continuous data (studies 4, 5, 13, 17, 26, 29, 31, 33, 45 and 52[*]), pooling did not alter the degree of observed heterogeneity and decreased the overall effect size. Subgroup analyses, once more, failed to remove all the statistical heterogeneity observed, but suggested that this effect might again be moderated by differences in baseline severity: a significant difference in favour of CBT was observed where severity was unspecified, but no difference was apparent in trials involving more severely depressed patients. The effect might also have been moderated by differences in the number of sessions offered: although less straightforward to interpret, trials involving fewer sessions appeared to demonstrate a greater difference in favour of CBT – an observation in the opposite direction from that in comparison 1.

[*] See appendix 6 for the references for each study.

Only five trials (providing six data sets) contributed dichotomous data on the number of non-symptomatic patients at follow-up (3 and 6 months and 1 year). Three of the trials were conducted in the USA (studies 5, 17 and 33*), one in the UK (study 49*) and one in Canada (study 28*). No statistical heterogeneity was apparent at any follow-up timepoint. Only one trial reported 6-month outcomes and suggested a significant difference in favour of CBT. The differences observed at post-treatment were not demonstrated at 3 months or 1 year and pooling of dichotomous data suggested no significant differences between the two groups. In the continuous data, five of the post-treatment outcome trials and one additional trial reported data contributing to the pooled mean differences, and all were conducted in the USA. Two trials contributed to the pooled mean change, both, again, conducted in the USA. No statistical heterogeneity was apparent at any timepoint in either of the two sets of data. Pooling the data at each timepoint suggested that those treated with CBT exhibited significantly fewer symptoms than the comparison groups at 3 months. Data could not be pooled at 6-month follow-up. There were no apparent differences between groups in symptom reduction from baseline to a follow-up of 3 months, although only small numbers of patients contributed to the analysis.

Although dropouts appeared to be slightly greater in the CBT group, there appeared to be no significant differences in dropout rate between those receiving CBT and those in the comparison groups. The number of dropouts was significantly lower in the CBT group in the trials involving patients with unspecified levels of depression. Summarising the available dropout data suggested a 21% dropout rate from the CBT group and 22% from the other three models, and it is unlikely that differential dropouts strongly influenced clinical outcomes.

In the recovery data, the observed difference between groups was more extreme in favour of CBT in those trials recruiting self-selected patients and volunteers than in those trials involving clinic-attenders and outpatients (which no longer demonstrated a significant difference between interventions), although statistical heterogeneity was still apparent in both categories. A similar effect was observed in the dropout data, but not in the mean differences data. The overall quality score of trials appeared to have a considerable effect on outcome for recovery and mean differences. Trials with lower scores demonstrated a pronounced and highly significant difference and higher-scoring trials demonstrated no significant differences. However, statistical heterogeneity was still apparent. Pooling dropouts according to overall quality score demonstrated no significant differences between groups in any category.

This comparison combined all available economic evaluations, but none was sufficiently large to allow robust testing of efficiency differences between psychotherapy and treatment as usual. However, each suggested that psychotherapy might have been superior with larger samples. Two of the four cost-effectiveness analyses were generally well designed and well conducted and provided tentative support for the hypothesis that psychotherapy is more efficient than usual care.

Individual therapy versus group therapy

Five out of the six trials contributing to the post-treatment recovery analysis were RCTs and one was a CCT (study 41*). They were fairly consistent in terms of quality and the number of sessions offered, although there was considerable variation in depressive symptomatology at recruitment. All used CT and BT, and three of the trials were conducted in the USA and three in the UK. No statistical heterogeneity was apparent. Pooling indicated that patients receiving individual therapy were significantly more likely than those receiving group therapy to improve to a degree where they were no longer considered to be clinically depressed.

In the post-treatment continuous data, only mean differences were reported. All but one (study 41*) of the eight trials were reportedly randomised, and all used formal CT, BT or CBT models. Five were conducted in the USA (studies 6, 41, 43, 59 and 63*) and three in the UK (studies 40, 47 and 55*). Again, no statistical heterogeneity was apparent. The results obtained by pooling dichotomous outcomes were supported by the findings from continuous data, and suggested that those treated with individual therapy exhibited significantly fewer symptoms post-treatment than those treated with group therapy.

Only three trials contributed dichotomous data on the numbers of non-symptomatic participants at follow-up (2 and 6 months). Two of the trials

* See appendix 6 for the references for each study.

were conducted in the USA (studies 6 and 63*) and one in the UK (study 55*). No statistical heterogeneity was apparent. The differences observed at post-treatment were not demonstrated at 2 or 6 months, because pooling of dichotomous data suggested no significant differences between the two groups.

Of the six trials providing continuous data at follow-up, all reported mean differences only at 1, 2, 3 and 6 months and 1 year. Three were conducted in the USA (studies 6, 43 and 63*) and three in the UK (studies 40, 47 and 55*). Statistical heterogeneity was apparent in the 6-month follow-up data. Pooling the data at 1, 2, 3 and 6 months suggested no significant differences between groups. Data at 1 year could not be pooled.

Although the number of dropouts appeared to be greater in group formats, there were no significant differences in dropouts between those receiving individual and group treatment. Summarising the available dropout data suggested 13% dropouts from individual therapies and 17% from group therapies and it is unlikely that differential dropouts strongly influenced clinical outcomes. Insufficient data prevented further sensitivity analyses being undertaken. No economic evidence was available for this comparison.

CT versus BT

Although 12 trials were identified as eligible for this comparison, only three had sufficient information for inclusion in the post-treatment recovery analysis. All three were reportedly randomised. Two were conducted in the USA (studies 23 and 29*) and one was conducted in Canada (study 50*). There was no evidence of statistical heterogeneity. One recent trial (study 23*) carried over 70% of the pooled OR weight. This was a well-conducted model-dismantling trial with a sample size of 50 patients or more in each arm (and the authors of this study admitted a bias towards CT, expressing surprise and disappointment at the equivocal outcomes produced by the cognitive and behavioural interventions). Pooling the data demonstrated no significant differences in post-treatment recovery between CT and BT.

Six trials provided continuous post-treatment data (mean differences only) and all were reportedly randomised. Five of the mean differences trials were conducted in the USA (studies 23, 29, 39, 53 and 63*) and one in Australia (study 60*). There was no evidence of statistical heterogeneity. The results obtained by pooling dichotomous outcomes were supported by the findings from continuous data. There were no significant differences between groups in terms of post-treatment symptomatology.

Only one trial contributed dichotomous data on the number of non-symptomatic patients at a follow-up of 6 months. In the continuous data, all six trials that provided post-treatment outcome data reported data on the mean differences at follow-up, three reporting at up to 2 months and three reporting at between 2 and 6 months. No statistical heterogeneity was apparent at any timepoint in either of the two sets of data. One trial (study 39*) that had recruited participants through media advertisements but involved people with severe depression, reported unusually high group means at follow-up. Pooling the data revealed no differences in symptom reduction from baseline between CT and BT at any timepoint.

There were no differences in dropouts between CT and BT. Summarising the available dropout data suggested 11% dropouts from CT models and 13% from BT models and it is unlikely that differential dropouts strongly influenced clinical outcomes. Insufficient data prevented further sensitivity analyses being undertaken for this comparison, and there was no economic evidence available for this comparison.

CBT versus treatment as usual

For this comparison, 29 trials were eligible although only 12 provided post-treatment recovery data. These trials were similar to those in comparison 1. However, a smaller number of patients were included because fewer psychotherapy arms were eligible from studies that compared more than one psychological treatment. Of the 12 trials contributing to the recovery analysis, all were reportedly RCTs. Six of the trials were conducted in the USA (studies 16, 25, 33–35 and 48*), four in the UK (studies 40, 45, 46 and 54*), one in Canada (study 50*) and one in Australia (study 58*). In contrast to comparison 1, despite all but one of the ORs favouring CBT, statistical heterogeneity was indicated in this outcome (suggesting that collapsing variants of psychotherapy arms within studies in comparison 1

* See appendix 6 for the references for each study.

reduced differences between groups). Removal of the one negative trial reduced statistical heterogeneity to a level that was non-significant. Pooling resulted in a highly significant OR, indicating that patients receiving variants of CBT were significantly more likely than those receiving treatment as usual to improve to a degree where they were no longer regarded as being clinically depressed. Subgroup analyses suggested that this effect was moderated by differences in baseline severity. A highly significant difference in favour of CBT was observed where severity was unspecified, and a slightly reduced, but still highly significant, difference in favour of CBT was seen in trials involving more severely depressed patients, although statistical heterogeneity was still apparent in the latter group. The overall effect did not appear to be moderated by differences in the number of sessions offered.

All but two (studies 7 and 51*) of the 20 trials contributing to the mean differences analysis and all of the five contributing to the mean change in symptoms analysis were reportedly randomised. Of the trials reporting mean differences, 14 were conducted in the USA (studies 3, 6, 7, 9, 14, 25, 33–35, 38, 43, 48, 51 and 53*), three in the UK (studies 40, 45 and 46*), two in Australia (studies 58 and 60*) and one in Canada (study 30*). Two of the trials reporting mean change in symptoms were conducted in the USA (studies 19 and 21*), two in the UK (studies 45 and 46*) and one in Australia (study 58*). Statistical heterogeneity was highly significant in the mean differences data, but no statistical heterogeneity was apparent in the mean change data. The results obtained by pooling the dichotomous data on whether participants were non-symptomatic were broadly supported by the findings from the continuous data. These suggested that those treated with CBT exhibited significantly fewer symptoms post-treatment than those on waiting lists or receiving treatment as usual, and that symptom reduction from baseline was greater in those receiving psychotherapy. Nine trials reported both recovery and mean differences data (studies 25, 33–35, 40, 45, 46, 48 and 58*) and it was possible to undertake a formal comparison between dichotomous and continuous data. Re-analysing the mean differences data from these trials slightly reduced the heterogeneity, but also reduced the effect size and widened the CI. However, in this analysis, dichotomous and continuous outcomes broadly supported one another. Subgroup analyses of all the trials, once more, failed to remove the statistical heterogeneity, but suggested that this effect might, again, be moderated by differences in baseline severity. A trend was observed demonstrating an increasing difference in favour of CBT with decreasing severity. The number of sessions offered appeared to make little difference to outcome.

Only four trials contributed dichotomous data on the number of participants who were non-symptomatic at follow-up (3 months). Three of the trials were conducted in the USA (studies 25, 35, and 48*) and one in the UK (study 54*). Statistical heterogeneity was significant primarily due to one trial (study 48*). The difference observed at post-treatment was not demonstrated at 3-month follow-up because pooling of dichotomous data suggested no significant difference between the two groups. In the continuous data, ten of the trials that provided post-treatment data also reported data contributing to the pooled mean differences at follow-up (1, 2, 3 and 6 months and 1 and 2 years). All but one trial (study 7*) were reportedly randomised and all but one (study 46* that was conducted in the UK) were conducted in the USA. Two trials contributed to the pooled mean change at follow-up (3 and 6 months and 1 year), one of which was conducted in the USA (study 25*) and one in the UK (study 46*). Statistical heterogeneity was apparent at several of the timepoints in the mean differences data, but not in the mean change data. Insufficient power prevented any firm conclusions being drawn from the mean differences data (significant differences were demonstrated at 1 and 3 months but not at 2 and 6 months), although the data suggested that those treated with CBT exhibited fewer symptoms than the comparison groups at 3 months. Data on follow-up at 1 and 2 years could not be pooled. However, the mean change data demonstrated statistically significant differences in symptom reduction from baseline in favour of CBT at 3 and 6 months, although data could not be pooled at 1 year.

Although dropouts appeared to be slightly greater in the CBT group than in those receiving treatment as usual, this did not reach statistical significance. Dropout rate was significantly greater in the CBT group in trials involving patients with severe depressive disorder, but not in those involving mild/moderate or unspecified levels of depression. There appeared to be no difference between CBT

* See appendix 6 for the references for each study.

and treatment as usual in those trials involving between one and six sessions, but a significantly greater number of dropouts from CBT was indicated in trials involving seven to 20 sessions. Summarising the available dropout data suggested 22% dropouts from CBT and only 11% from treatment as usual. Since dropouts were more likely to be unreported and we have assumed all missing data to be treatment failures, it is possible that we have under-estimated the benefits of CBT.

Although significant differences were still observed in all categories in the recovery and mean differences data, the effect in favour of CBT was greater in those trials recruiting self-selected patients and volunteers than in those involving clinic-attenders and outpatients. However, statistical heterogeneity was still apparent in both categories in the mean differences data. No differences between trials according to the source of recruitment were demonstrated in the dropout data. In the recovery and mean differences analyses, trials with lower overall quality scores generally demonstrated greater effect sizes in favour of CBT than higher-scoring trials. A significant difference in dropouts was only suggested in the lower-quality trials in favour of treatment as usual. A funnel plot drawn from data from this comparison again suggested evidence of bias by demonstrating the absence of small- and medium-sized negative trials. Thus, once more, the importance of interpreting the findings with caution cannot be over-emphasised.

The only available cost-effectiveness analysis (linked to an excluded study by Katon and colleagues; see appendix 7 for the reference for this study) concluded that CBT offered within a collaborative care model for MDD was more costly than treatment as usual, but achieved greater success. This translated into a modest cost-effectiveness advantage. The same intervention for minor depression was more costly but not more cost-effective than treatment as usual.

CBT versus IPT

The data in this comparison consisted of just two RCTs, which had been extensively reported and discussed in the literature. The USA National Institute of Mental Health (NIMH) Collaborative Research into Depression Program (study 13*) generated 11 papers from its trial, and the UK-based Second Sheffield Psychotherapy Project (study 49*) produced ten papers. Furthermore, both trials caused considerable debate in the literature regarding differing aspects of methodology. The NIMH trial reportedly used therapists less experienced in the CBT condition than in the IPT condition.[62] It was also criticised for failing to stratify severity of depression at baseline. Some commentators suggested that the Second Sheffield Psychotherapy Project trial lacked statistical power to detect treatment differences. No statistical heterogeneity was apparent and no significant differences were demonstrated.

The same trials provided continuous data at post-treatment, one reporting mean differences (study 13*) and one reporting mean change (study 49 (both eight- and 16-session arms)*). Again, no statistical heterogeneity was apparent. The analyses suggested no differences between CBT and IPT, supporting the findings from the recovery data. The UK trial reported dichotomous outcomes on whether participants were non-symptomatic at 1-year follow-up and demonstrated no difference between the two therapies. No continuous follow-up data were available.

Although dropouts in the individual trials appeared to be greater in CBT, pooling the data demonstrated no significant difference in dropout rate between the two therapies. A total of 27% dropped out of CBT compared with 20% from IPT. Since dropouts were high in both groups, this could threaten generalisability of findings from this comparison. Insufficient data prevented further sensitivity analyses being undertaken, and no economic evidence was available for this comparison.

CBT versus PDT

Six trials (contributing seven sets of data) were included in the post-treatment recovery analysis. Five of the six were RCTs (studies 4, 11, 17, 28 and 49*) and one was a CCT (study 26*). All six used diagnostic inclusion criteria to establish MDD and conducted the psychotherapy over more than ten sessions. Four of the trials were conducted in the USA (studies 4, 11, 17 and 26*), one in the UK (study 49*) and one in Canada (study 28*). It is noteworthy that three of the six trials presented the PDT arm as a control condition (studies 4, 11 and 26*), which may have been suggestive of bias on the part of the researchers towards the CBT condition in more than half of the trials. No statistical heterogeneity was apparent, although the only CCT in the

* See appendix 6 for the references for each study.

group reported a non-significant OR in favour of PDT (study 26*). Pooling suggested that patients receiving CBT were significantly more likely to improve to a degree where they were no longer regarded as being clinically depressed than those receiving PDT.

Four trials (contributing five data sets) provided continuous data at post-treatment: three reported mean differences data (studies 4, 17 and 26*) and two contributed mean change data (studies 17, 49 (eight-session arm) and 49 (16-session arm)*) Three were RCTs (studies 4, 17 and 49*) and one was a CCT (study 26*). Three were conducted in the USA (studies 4, 17 and 26*) and one in the UK (study 49*). Significant statistical heterogeneity was observed in the mean differences data, where, again, study 26* reported an extreme and significant effect in the opposite direction from the other two trials. No statistical heterogeneity was apparent in the mean change data, and both sets of data demonstrated no significant differences post-treatment between CBT and PDT.

Three trials contributing four sets of data provided dichotomous data on the number of non-symptomatic patients at follow-up (3 months and 1 year). One trial was conducted in the USA (study 17*), one in the UK (study 49*) and one in Canada (study 28*). No statistical heterogeneity was apparent at either timepoint. The differences in recovery observed at post-treatment were not demonstrated at 3 months and 1 year: pooling of dichotomous data suggested no significant difference between CBT and PDT. Two trials (both conducted in the USA) provided continuous data at follow-up, one providing mean differences data only (study 26*) and one providing both mean differences and mean change from baseline (study 17*) at follow-up (both 3 months). No statistical heterogeneity was apparent, and pooling the data for both outcomes demonstrated no differences between groups.

Although the overall number of dropouts appeared to be greater from PDT than from CBT, pooling did not result in a significant difference between groups. Insufficient data prevented further sensitivity analyses being undertaken. No economic evidence was available for this comparison.

CBT versus ST

Seventeen trials were eligible for inclusion in this part of the review, but only ten provided sufficient post-treatment recovery data for meta-analysis. Of these, eight were RCTs (studies 5, 16, 28, 29, 31, 33, 45 and 50*) and two were CCTs (study 52 and 56*). Seven of the trials were conducted in the USA (studies 5, 16, 29, 31, 33, 52 and 56*), one in the UK (study 45*) and two in Canada (studies 28 and 50*). Statistical heterogeneity was highly significant in these data. Several of the trials (studies 33, 52 and 56*) reported extreme ORs (≥ 15) and a single relatively large trial suggested a significant finding in the opposite direction (study 45*). The analysis indicated that patients receiving variants of CBT were significantly more likely to improve to a degree where they were no longer considered clinically depressed than those receiving ST. Subgroup analyses removed the observed statistical heterogeneity and suggested that this effect might be moderated by differences in baseline severity: a highly significant difference in favour of CBT was observed where severity was unspecified, but a non-significant difference was apparent in trials involving more severely depressed patients. The overall effect also appeared to be moderated by the number of sessions offered: the greatest difference in favour of CBT was observed in trials involving fewer sessions and those involving more sessions demonstrated no significant difference. This is, again, an observation in the opposite direction to that in comparison 1, but in the same direction as in comparison 2, although, once more, statistical heterogeneity was still apparent.

All but two (studies 36 and 52*) of the nine trials contributing to the mean differences post-treatment analysis and all but one (study 36*) of the three contributing to the mean change analysis were reportedly randomised. Seven of the trials reporting mean differences were conducted in the USA (studies 5, 22, 29, 31, 33, 36 and 52*) one in the UK (study 45*) and one in Canada (study 30*). Two of the trials reporting mean change data were conducted in the USA (studies 5 and 36*) and one in the UK (study 45*). Statistical heterogeneity was highly significant in the mean differences data. Although reportedly measuring outcome on the same scale, several trials (studies 30, 36 and 52*) reported unexpectedly high group means. Although no statistical heterogeneity was apparent in the mean change data, one of these trials reporting high mean differences also appeared in this analysis (study 36*) and again reported considerable mean changes in both groups. This was quite different from the

* See appendix 6 for the references for each study.

other trials in the group. Findings from the mean differences analysis supported the results obtained by pooling dichotomous outcomes, which suggested that those treated with CBT exhibited significantly fewer symptoms post-treatment than those receiving ST. Although there appeared to be a trend in favour of variants of CBT in the mean change data, there were no significant differences between groups in terms of symptom reduction from baseline. Subgroup analyses once more failed to remove all the statistical heterogeneity observed, but suggested that this effect might again be moderated by differences in baseline severity. Again, significant differences in favour of CBT were observed where severity was unspecified, but no difference was apparent in trials involving more severely depressed patients. The effect might also have been moderated by differences in the number of sessions offered: although less straightforward to interpret, trials offering fewer sessions appeared to demonstrate a greater difference in favour of CBT (an observation in the opposite direction from that in comparison 1, but in the same direction as comparison 2).

Only three trials contributed dichotomous data at follow-up (3 and 6 months). Two of the trials were conducted in the USA (studies 5 and 33*) and one in Canada (study 28*). No statistical heterogeneity was apparent, although only a small number of trials were included. Only one trial reported 6-month outcomes and suggested a significant difference in favour of CBT. The differences observed at post-treatment were not demonstrated at 3-month follow-up: pooling of dichotomous data suggested no significant difference between the two groups. In the continuous data, three of the trials reporting post-treatment outcomes provided data contributing to the pooled mean differences at follow-up, all of which were conducted in the USA. Only one trial contributed to the pooled mean change at 3-month follow-up. No statistical heterogeneity was apparent at any timepoint. Pooling the continuous follow-up data suggested that those treated with CBT exhibited significantly fewer symptoms than those treated with ST at 3-month follow-up, however, data could not be pooled at 6-month follow-up.

Although dropouts during treatment appeared to be slightly greater in the CBT group, there appeared to be no significant differences in dropouts between those receiving CBT and those receiving ST. Subgroup analyses suggested that this effect might, again, be moderated by differences in baseline severity: dropouts from CBT might have been greater in trials involving more severely depressed patients and less in trials where severity was unspecified, but the difference did not reach statistical significance in either case due to a lack of statistical power.

In the recovery data, the observed difference between groups was extreme and highly significant in favour of CBT in those trials recruiting self-selected patients and volunteers, but no difference between groups was demonstrated on pooling those trials including clinic-attenders and out-patients. However, statistical heterogeneity was still apparent in both categories. No significant differences were observed between groups according to the source of recruitment in the continuous or dropout data. The overall quality score of the trials appeared to have a considerable effect on recovery and mean differences, with lower-scoring trials demonstrating a pronounced and highly significant difference and higher-scoring trials demonstrating no significant differences.
A similar effect was observed in the dropout data, although pooling according to overall quality score demonstrated no significant differences between groups in either category.
No economic evidence was available for this comparison.

IPT versus ST

This analysis included only one small, USA-based, randomised pharmacotherapy/psychotherapy combination trial, with the pharmacotherapy conditions excluded. There appeared to be no significant differences in recovery or dropouts between IPT and ST. No continuous data or follow-up data were available, and there was no economic evidence for this comparison.

IPT versus treatment as usual

This comparison consisted of just one USA-based RCT (study 44*), involving a large number of patients with arms of more than 90 patients in both conditions included in the review. This trial suggested that recovery was significantly more likely for those receiving IPT than for those receiving treatment as usual and that this difference was maintained at 8-month follow-up. The continuous mean differences and mean change data from this trial supported this finding, again, indicating a significant difference in favour

* See appendix 6 for the references for each study.

of IPT. However, it should be noted that attrition was reported to be high from the IPT arm and that 45% of those allocated to treatment as usual were given antidepressant therapy during the course of the trial. IPT was shown by the only available economic analysis to be more cost-effective than usual care only if delivered by non-psychiatrists.[60]

PDT versus ST

One Canadian RCT provided data for this comparison. The ST under evaluation was relaxation therapy, which, although behavioural in theoretical origin, was quite explicitly used in this study as an attention-placebo control. The trial indicated no differences in recovery post-treatment or 3-month follow-up, or in the number of dropouts. It should be noted, however, that the randomisation procedure in this trial was questionable because dropouts were apparently replaced to maintain group sizes, thereby introducing a probable bias. No economic evidence was available for this comparison.

ST versus treatment as usual

All four trials contributing to the post-treatment recovery analysis were reportedly RCTs (studies 16, 33, 45 and 50*). Two were conducted in the USA (studies 16 and 33*), one in the UK (study 45*) and one in Canada (study 50*). No statistical heterogeneity was apparent. However, it should be noted that one trial in this comparison (study 45*) accounted for 65% of the weight in the pooled analysis. Although not apparent in this comparison, this trial repeatedly demonstrated very different results from other trials in the review, possibly due to differences in the application of therapeutic techniques, and perhaps caution is required in interpreting the results of this comparison. The pooled result suggested that patients receiving ST were significantly more likely than those receiving treatment as usual to improve to a degree where they were no longer regarded as being clinically depressed.

Four RCTs provided post-treatment mean differences data and two of these also provided mean change data. Two were conducted in the USA (studies 33 and 57*), one in the UK (study 45*) and one in Canada (study 30*). No statistical heterogeneity was apparent. The results obtained by pooling the dichotomous recovery data were only partially supported by the findings from the continuous data. Pooling the latter suggested that those treated with ST exhibited significantly fewer symptoms post-treatment than those receiving treatment as usual, yet no difference in mean change from baseline was demonstrated between the groups.

No dichotomous data on whether participants were non-symptomatic at follow-up were available. Two trials (both conducted in the USA) provided continuous follow-up data, one providing both mean differences and mean change data at 2 months and one providing mean differences data at 6 months. No differences between groups were demonstrated.

No significant differences between groups in terms of dropouts during treatment were demonstrated. Summarising the available dropout data suggested 22% dropouts from ST and only 11% from treatment as usual. Since dropouts were more likely to be unreported, and we have assumed all missing data to be treatment failures, it is possible that we have slightly under-estimated the benefits of psychotherapy. Insufficient data prevented further sensitivity analyses being undertaken for this comparison.

Only one trial included an economic component (linked to an excluded study by Katon and colleagues and referred to earlier in comparison 5; see appendix 7 for the reference for this study) and concluded that ST offered within a collaborative care model was more costly but achieved greater success than treatment as usual.

Robustness of the review

The searches for this review were thorough and comprehensive, and sought to identify published and unpublished trials. Even so, all five funnel plots indicated that small negative trials might have been omitted. It is, therefore, likely that the positive treatment effects of smaller studies have been exaggerated. It is also important to note that statistical heterogeneity was a common feature in a number of comparisons for several outcomes, and that CIs tended to be wide. Potential biases may have been the result of a number of different factors.

(1) The absence of smaller studies with no statistically significant effects could result in a pooled over-estimate of treatment effect. Publication bias and other reporting biases

* See appendix 6 for the references for each study.

are very likely explanations for possible missing studies – all of the trials in this review were published, in English and cited elsewhere, and many had resulted in multiple publications. All of these factors would have increased their likelihood of identification and inclusion in the review. Furthermore, as demonstrated in the sensitivity analyses, trials of lower methodological quality (although not necessarily the smaller studies) generally demonstrated larger effects, which might have possibly increased their chances of publication. The potential for selection bias needs to be considered when interpreting the findings of this review.

(2) For clinical outcomes, we assumed that all unreported data in identified trials represented treatment failures. Outcomes from dropouts are more likely to be unreported and it is possible, therefore, that we have over-estimated treatment failures in both groups.

(3) The findings from heterogeneity and sensitivity analyses also suggest that differences in probability of recovery may also partially explain the findings. Studies involving volunteers and patients with depression of unspecified severity tended to be smaller. Furthermore, there was some evidence to suggest that earlier trials had been smaller than later ones, and the effects of changes in both the intervention and comparator groups over time cannot be ruled out. Thus, the results may reflect true heterogeneity amongst the trial groupings.

(4) Some evidence from trial quality rating and from sensitivity analyses suggested that lower-quality studies tended to be smaller and resulted in larger effects. It is not possible to know whether those studies not contributing to the meta-analysis were small negative trials, since a number of them did not specify even the overall sample size.

(5) Pooled sample sizes remained inadequate in many of the outcome analyses and, therefore, the power to detect potentially significant differences between the different forms of psychotherapy was limited.

Thus, the potential influence of chance and bias from a number of different sources limits confidence in the overall results and increases caution in our interpretation of these findings.

Generalisability of the trials contributing to the review findings

Socio-economic characteristics of patients

Many sample groups in USA-based trials consisted of well-educated subjects of higher socio-economic levels. Fifteen trials included female patients only (studies 4, 8–10, 16, 19, 21, 24, 26, 27, 36–39 and 62[*]). Sixteen trials used volunteers on university campuses, with younger than average participants of higher educational and socio-economic status. One small cluster of university-based trials further limited their study populations to all students in the same year enrolled on psychology courses. In contrast, however, three trials selected sample groups from a low-income population only. The generalisability of these trials to patients presenting in primary care settings and to settings within the UK, therefore, remains questionable.

Severity of depression

Patients in the included trials showed considerable variation in symptoms of depression, with diagnostic inclusion criteria ranging from negative mood state to MDD. Some trials failed to conduct a diagnostic interview at recruitment, relying solely on a self-report instrument to measure levels of depressive symptoms. Not surprisingly, clinical heterogeneity was indicated. Of course, it is possible that the wide-ranging presentation of depressive symptoms represented in these studies may be a reflection of the variation in depression encountered by healthcare professionals in primary care settings and, therefore, could be regarded as reasonably generalisable in terms of clinical presentation for treatment.

Country bias

A considerable proportion of trials in the review (50 of 63) were conducted in the USA, and just 13 studies were carried out in other countries, of which seven were UK-based trials. One of the notable differences between USA studies and those from other countries appeared to be the clinical background and qualifications of the therapists employed. In the USA trials, the largest group of therapists were 'advanced' (often PhD) clinical psychology or counselling psychology students. This group of professionals in training may be representative of therapists in the USA, but it is difficult to be certain of the extent to which their

[*] See appendix 6 for the references for each study.

experience and qualifications might generalise to those of therapists in the UK primary care setting.

Motivation of participants

More than half the trials used volunteer populations by obtaining sample groups through local radio or newspaper advertisements, and sometimes even offering small cash payments to subjects who agreed to take part in studies. Patients who volunteer to participate in intervention studies are a self-selected group who tend to experience generally lower morbidity rates than those who do not take part,[63] and whose motivation for treatment and attrition rates may differ from the entire experimental population and from the general patient population.

Inclusion of trials that had identified and recruited patients in a more systematic way through outpatient clinics and referrals generally resulted in slightly lower effect sizes than was obtained by grouping all of the trials. These trials were likely to have involved a more homogeneous group of participants in terms of pathways to care, patient motivation, severity, chronicity, previous treatment, etc., but were also more likely to have involved patients that clinicians would have expected to benefit from treatment. There are several possible explanations for the observed differences in outcomes between these trials and those involving self-selected volunteers and responders to advertisements. Perhaps the single most important of these is the probable difference in patient motivation, both in their participation in the trial and their commitment and determination in therapy. It should be noted that the generalisability of trials involving self-selected subjects to clinical practice might be poor. Nevertheless, it is likely that this self-selected group may have included some patients whose severity of depression was more similar to that of patients in primary care, as well as some whose depression remained undetected by the general practitioner. It is, therefore, important to note that, in all but one case, the pooled outcomes in this group of trials were as good as or better than those involving clinic attenders and outpatients.

Length of treatment and type of therapy

It could be suggested that the time-limited design of the psychological treatments investigated may have favoured goal-orientated structured psychotherapy models, which strive towards positive action from the outset of treatment. In contrast, the psychodynamic approach posits the expectation that as clients explore and confront their difficulties, their symptoms may become more severe before they begin to improve. It may be, therefore, that a greater number of sessions are required to work through this process in order for a positive outcome to be achieved. However, it should be stressed that the search strategy employed for this review failed to identify a single controlled trial evaluating the effectiveness of PDT for the treatment of depression over more than 20 sessions, and, given the current absence of relevant studies in the literature, it would not be possible to test that hypothesis. It should also be stressed that PDT and client-centred therapy were mostly non-manualised and non-standardised, which would limit their systematic application to clinical practice.

Long-term outcomes

Long-term outcomes were not measured in two-thirds of trials, and those trials that did provide long-term follow-up assessments of more than 6 months tended to suffer high attrition rates between post-treatment interviews and final follow-up.

Adverse events and tolerability

Generally, reporting of adverse effects resulting from psychological treatments was poor. Two pharmacotherapy/psychotherapy combination trials reported withdrawing patients from psychotherapy treatment due to an observed deterioration of symptoms. The author of one other trial reported adverse events in two patients receiving psychotherapy who carried out acts of deliberate self-harm whilst participating in the trial (study 36*), although it is unclear whether these were directly associated with the trial. No other trial appeared to consider possible adverse consequences of psychological treatment, nor were such complications suggested as a cause of attrition from trials. Indeed, reasons for patient dropouts were infrequently investigated or reported by authors. Dropout outcomes were originally summarised to provide an indication of tolerability. However, it is also worth noting the potential threat to generalisability of the review that might be introduced by high numbers of dropouts from trials. For the purposes of this

* See appendix 6 for the references for each study.

review, dropouts were conservatively regarded as 'treatment failures'. However, there is no way of knowing from the reports of most trials how many patients actually ceased to attend sessions due to an intolerable increase in their symptoms. It is not possible, therefore, to report on potential drawbacks or negative outcomes of psychotherapy treatment.

Economic evaluation

Out of a total of 63 trials in the review, just five included a cost-effectiveness component, demonstrating an overall paucity of economic data available for analysis and interpretation. The trials that provided economic evaluations showed considerable variation in the quality and scope of their cost measurements. Three of the trials were conducted in the USA, which made it difficult to judge the relevance of the economic findings because cost findings would have been sensitive to the healthcare system within which brief psychological interventions were delivered. Although the USA-based economic evaluations were well-conducted studies, it should be emphasised that the confidence with which those economic findings might be applicable to the UK would be lower than for clinical findings.

When investigating the cost-effectiveness dimension of a trial, it is of particular importance to have outcomes with long-term follow-up because the cost-offsetting effects may only be revealed over time. As noted earlier, only 14 trials (22%) reported follow-up dichotomous or continuous outcomes in terms of depressive symptoms, and just nine studies with sufficient data for inclusion in the meta-analysis conducted final follow-up assessments more than 6 months after the completion of psychotherapy. It was, therefore, not possible to conduct a long-term economic evaluation for the purposes of this review, and this should be regarded as another limitation.

Statistical power is also a particular problem in economic evaluations, because some events with important economic implications (such as inpatient admissions) may rarely happen but are very expensive when they do occur. The two well-conducted economic evaluations included in the review collected data on inpatient service utilisation, but excluded the costs from their analyses because of the skew in the data and the inadequacy of sample sizes. The other evaluations with an economic component had even smaller sample sizes. Overall, therefore, the review lacks adequate statistical power to conduct a meaningful cost-effectiveness analysis.

Chapter 5
Conclusions

Summary

Comparisons between treatments

Comparing any variant of psychotherapy with treatment as usual or waiting lists suggested patients receiving psychotherapy were significantly more likely to improve to a degree where they were no longer regarded as being clinically depressed. They also exhibited significantly fewer symptoms post-treatment and experienced significantly greater symptom reduction from baseline. Follow-up data for different timepoints were sparse, but the available evidence did suggest that the psychotherapy group continued to exhibit significantly fewer symptoms at follow-up periods of up to 9 months. The risk of treatment discontinuation in those receiving psychotherapy was no greater than in those receiving treatment as usual. Economic evidence provided tentative support for the hypothesis that psychotherapy was more efficient than usual care.

Trials of CBT, CT or BT versus PDT, IPT or ST suggested that patients receiving any of the cognitive behavioural group of therapies were significantly more likely to improve to a degree where they were no longer regarded as being clinically depressed. However, there were no differences between groups in post-treatment symptoms, symptom reduction from baseline or dropouts during treatment. It was not possible to draw any firm conclusions from the limited follow-up and economic data available.

Comparing individual with group therapy formats suggested patients receiving individual therapies were significantly more likely to improve to a degree where they were no longer regarded as being clinically depressed and exhibited significantly fewer symptoms post-treatment. No differences in dropouts between groups were demonstrated. Again, it was not possible to draw any firm conclusions from the limited follow-up and economic data available. (It should be noted that all the trials identified in this comparison used cognitive behavioural interventions.)

In trials comparing CTs and BTs, no differences in post-treatment recovery and symptoms, symptom reduction from baseline or dropouts were demonstrated. Only limited follow-up data were available, and no economic data were identified.

Comparing any variant of CBT with waiting lists or treatment as usual suggested patients receiving variants of CBT were significantly more likely to improve to a degree where they were no longer considered clinically depressed, exhibited significantly fewer symptoms post-treatment and experienced significantly greater symptom reduction from baseline. Follow-up data for different timepoints were sparse, but the available evidence did suggest that the CBT group continued to exhibit significantly fewer symptoms at follow-up periods of up to 6 months. The risk of treatment discontinuation in those receiving variants of CBT was no greater than in those receiving treatment as usual. In addition, the economic evidence suggested a modest cost-effectiveness advantage in favour of CBT.

The evidence comparing variants of CBT with IPT was too limited to draw any conclusions, although it did suggest that there were no differences in post-treatment recovery and dropouts during treatment between the two types of therapy. No economic evidence was available.

Comparing any variant of CBT with PDT suggested that patients receiving variants of CBT were significantly more likely to improve to a degree where they were no longer regarded as being clinically depressed, although no group differences in post-treatment symptoms, symptom reduction from baseline or dropouts during treatment were suggested. Only limited follow-up data were available, and no economic data were identified.

Comparing any variant of CBT with ST suggested that patients receiving variants of CBT were significantly more likely to improve to a degree where they were no longer considered clinically depressed and exhibited significantly fewer symptoms post-treatment. No group differences in symptom reduction from baseline or dropouts during treatment were suggested. Only limited follow-up data and no economic data were available.

Trials of IPT versus ST, IPT versus treatment as usual and PDT versus ST all yielded inadequate data upon which to base any firm conclusions.

Comparing ST with treatment as usual suggested patients receiving ST were significantly more likely to improve to a degree where they were no longer regarded as being clinically depressed and exhibited significantly fewer symptoms post-treatment than those receiving usual treatment. No group differences in symptom reduction from baseline or dropouts during treatment were suggested. Minimal follow-up data were available, although limited economic data suggested a cost-effectiveness advantage in favour of ST.

Whilst the data used in this review suggest an overall benefit of psychotherapy, and, in particular, CBT techniques, a number of caveats limit our confidence in these findings.

Internal validity

Low scores on internal validity items and inadequate reporting of methodology were a feature of the majority of the included trials, the overall quality of which were low. Only ten trials reporting the use of randomisation described their allocation procedure, and the possibility of investigator bias in the allocation of patients cannot be ignored. About one-third of the trials employed a waiting-list arm as a comparator, and it is possible that this could have influenced the size of the observed treatment effect by discouraging symptomatic improvement during the course of the trial in patients allocated to this intervention. In the majority of the trials, the authors had developed, or were closely associated with, the active therapy under evaluation. This may have been a potential for investigator bias. Furthermore, PDT and client-centred therapy tended to be regarded and utilised by investigators as attention-placebo conditions, again suggesting possible investigator bias. The implicit and explicit use of both antidepressants and concurrent non-randomised psychotherapy in a number of trials also limits confidence in the review findings, since it is uncertain that any observed treatment effect was achieved through the psychotherapeutic intervention alone. Finally, nearly 40% of the trials failed to monitor therapist adherence to the psychotherapeutic technique under evaluation, and it is not certain that the therapists consistently applied the models appropriately.

Low sample size was a feature of most of the trials. Only three trials were likely to have had adequate power to detect a real treatment difference. This is a problem of some importance in trials comparing two potentially 'active' treatments where the differences between groups might be expected to be relatively small. Additionally, more than half the trials excluded patients initially randomised who did not commence treatment or who later dropped out, limiting comparability between groups. Furthermore, despite the fact that people with depression are disabled in many spheres of activity, broader measures of outcome, such as quality of life, were rarely a feature of these trials, and the focus on one measure of clinical improvement is far from appropriate in patients with this disorder.

Robustness of the review

Despite the extensive searches undertaken, all 63 trials identified were published and were in English. In addition, many had resulted in multiple publications, thereby increasing their chances of inclusion in the review. Of these, 12 provided insufficient data for inclusion in the meta-analysis. The interpretation of the findings of this review is further limited by the identification of probable biases as indicated in the funnel plots and the heterogeneity and sensitivity analyses.

Although it is essential to be cautious in interpreting the findings of subgroup analyses, several consistent and recurring observations resulted from these. They suggested that the effects of treatment were likely to have been moderated by the baseline severity of depression in the trial participants, the recruitment setting and possibly the number of sessions provided. Trials of patients without a formal diagnosis and those involving volunteers consistently demonstrated larger treatment effects, and there was some evidence suggesting a similar effect in those trials of lower methodological quality. These observations suggest that the potential influence of bias on this review cannot be under-estimated, whilst the statistical heterogeneity and limited statistical power evident in many of the comparisons increases the likelihood of chance findings.

Generalisability

The generalisability of the review findings to UK primary care settings is questionable. Of the 63 trials included in the review, 52 were conducted in the USA. In the USA trials, the subjects tended to be well educated and of higher socio-economic status than might be found in depressed out-patients in the UK. A significant proportion of the trials included only females, and half the trials used volunteers. Notable differences in therapist background and qualifications were recorded between the USA studies and those from other countries, and doubt remains about the generalisability of the therapist intervention to UK clinical practice. More than half the trials used self-selected volunteer participants, many of whom

had no formal diagnosis of depression. Notable differences in outcomes between trials of volunteers and those involving outpatient clinic populations were suggested, and it is possible that varying levels of motivation for and commitment to treatment partially accounted for this finding. Considerable differences in depressive symptoms were also noted in the trials in this review. (However, whilst this clinical heterogeneity had a notable impact on the outcomes reported, it might also be argued that the range of reported severity improves the generalisability of the review findings to UK primary care settings).

The applicability of the interventions used in the trials reviewed here is also doubtful. Details of the PDT and client-centred therapy were often not provided and they were mostly non-manualised and non-standardised, which limits their utility in clinical practice. The use of both antidepressants and concurrent non-randomised psychotherapy in a number of trials may also fail to mirror the clinical circumstances in which patients in UK general practice are treated.

Almost every trial failed to report the existence of any adverse effects of therapy. Such effects were not considered as a potential cause of attrition from trials. Furthermore, reported attrition rates appeared to be considerably lower than might have been expected, and there is no way of knowing how many patients ceased to attend psychotherapy sessions due to an intolerable increase in their symptoms. Thus, conclusions cannot be drawn about potential adverse effects of psychotherapeutic treatments. In addition, due to the lack of long-term follow-up in two-thirds of the trials, as well as high attrition rates in those that did report follow-up, it is not possible to comment on the potential longer-term benefits of brief psychological treatments.

Economic evaluation

Conclusions from the synthesis of economic data were limited by the settings and countries in which they had been conducted, the notable lack of statistical power in the trials and the paucity of reports involving long-term follow-up. The review lacks adequate statistical power to conduct a meaningful secondary cost-effectiveness analysis.

Implications of the review for healthcare

Overall, the implications of this review relate more to USA settings than to the UK and other countries. In terms of the implications for clinical practice, based on the best available evidence, it would appear that some forms of brief psychotherapy, particularly those derived from cognitive/behavioural models, are of benefit in the treatment of people with depression who are being managed outside hospital settings.

Individual therapy may be more effective than group therapy. However, little evidence was found that explicitly examined the efficacy of individual versus group formats, and that which was identified used cognitive or behavioural interventions. Therefore, little can be said about the implications for individual versus group therapy in general.

A consistent finding from subgroup analyses suggested that the efficacy of psychological treatments might be influenced by baseline severity, the methods used to identify patients and, possibly, the number of sessions offered. However, the considerable caution required to interpret the findings from subgroup analyses cannot be overemphasised, and it is not possible to state conclusively how these factors might influence outcome based on the available evidence from these trials. Thus, other than to say that these factors are likely to be important, little can be provided in the way of guidance for clinical practice.

Similarly, little can be concluded about the potential impact of socio-demographic characteristics of patients, the specific effects of client motivation and therapeutic alliance, any potential adverse events associated with psychological treatments, the short- and long-term outcomes of psychological treatments, the differential effects of alternative models (particularly PDT and client-centred therapy) or the immediate and long-term economic consequences attached to the provision of psychological treatments in primary care settings.

Implications for research

The review has clear implications for future research priorities, as well as highlighting aspects of psychotherapy trial design and reporting requiring particular attention in future RCTs.

Firstly, it is important to note that although a difference between alternative treatments was often suggested in most of the comparisons, these frequently failed to reach a level of statistical significance that would support any firm conclusion due to a lack of power. Thus, the

review highlights the need for further trials of all types of psychological treatments in primary care settings involving appropriately recruited representative samples of participants whose disorders have been recognised and who meet the recognised diagnostic criteria for depressive disorder.

This latter point is particularly important. Given that the presentation of depression in primary care settings varies widely in terms of levels of severity and chronicity, aetiology and many other factors and that many depressed patients are likely to recover in the absence of treatment,[64] it is likely that psychological treatments are more appropriate for some patients than for others. It is important that new trials establish the actual degree of improvement that might be expected in patients with different levels of severity in view of:

(a) the repeated observations from the subgroup analyses on the influence of baseline severity, which strongly suggested that the treatment effect was smaller (even non-existent sometimes) than the overall effect for patients with verified MDD but greater for those who did not necessarily meet the criteria for a formal diagnosis of depression
(b) the fact that little is known about the potential adverse consequences and cost-effectiveness of different psychological treatments.

Future research needs to demonstrate whether psychological interventions are appropriate in all cases.

To be of any assistance in informing policy and practice, future trials should be adequately powered, involve longer follow-up periods, properly monitor adherence to therapeutic technique and, where non-randomised concomitant treatments are allowed, record and allow for these in the interpretation of findings. Once again, the need for adequately powered high-quality cost data in trials of this nature, particularly those with longer-term follow-up, and the importance of incorporating outcomes that measure the broader impact of these treatments cannot be over-emphasised.

Secondly, in light of the therapeutic techniques used in UK clinical practice, there is a particularly pressing need for trials investigating the effectiveness of PDT or client-centred therapy using manualised/standardised techniques in both brief and longer-term formats, and involving longer-term follow-up.

Thirdly, a much better evidence base from RCTs is required on the use of individual and group formats examining both immediate and long-term outcomes, again, incorporating the collection of good-quality cost information and comparing a range of different psychological treatments, particularly alternatives to CBT.

Acknowledgements

The authors would like to thank Dr Stuart Logan at the Systematic Reviews Training Unit, Institute of Child Health for his advice and Rosie Mundt-Leach, Vanessa Cooper and Noel Fitzgerald for their contributions to handsearching the journals. They are very grateful to Dr Michael Barkham and Professor Eugene Paykel who sent copies of papers about publications in press, and to the following people (listed alphabetically) who helpfully responded to the letter seeking further information on their trials: Professor L Beutler (who supplied new data), Professor C Conoley, Dr L Covi, Ms D Ellefson (who responded on behalf of Dr A Fazio, deceased), Dr D Gallagher-Thompson (who supplied new data), Professor L Greenberg, Dr J Greist, Dr P Hayman, Dr W Katon (who supplied new data), Professor D Kavenagh (who supplied new data), Professor P McLean, Professor R Neimeyer, Dr T Pace, Dr M Padfield, Dr L Range, Professor L Rehm (who supplied new data), Dr R Rosenthal (who responded on behalf of Dr T Rosenthal, deceased), Dr A Rush, Dr C Schulberg (who supplied new data), Dr A Scott (who supplied new data), Dr C Scott (who supplied new data), Dr M Scott (who supplied new data), Professor D Shapiro (who supplied new data), Dr J Teasdale, Dr M Weissman and Dr T Zeiss.

The authors are also indebted to the referees for their attention to the report and the quality of their comments. The views expressed are those of the authors, who are also responsible for any errors.

References

1. Paykel ES, Priest RG. Recognition and management of depression in general practice: consensus statement. *BMJ* 1992;**305**:1198–202.
2. Hays RD, Wells KB, Sherbourne CD, Rogers W, Spritzer K. Functioning and well-being outcome of patients with depression compared with chronic general medical illnesses. *Arch Gen Psychiatry* 1995;**52**:11–19.
3. Hotopf M, Lewis G, Normand C. The treatment of depression: evaluation and cost-effectiveness. PHP Departmental Publication. London: LSHTM; 1996.
4. Mann A. Depression and anxiety in primary care: epidemiological evidence. In: Jenkins R, Newton J, Young R, editors. The prevention of depression and anxiety. London: HMSO; 1992.
5. Gotlib IH, Colby CA. Treatment of depression – an interpersonal systems approach. New York: Pergamon Press; 1987.
6. Freeling P, Tylee A. Depression in general practice. In: Handbook of affective disorders. Edinburgh: Churchill Livingstone; 1992.
7. Blacker C, Clare A. Depressive disorder in primary care. *Br J Psychiatry* 1987;**150**:3–51.
8. Kind P, Sorensen J. The cost of depression. *Int Clin Psychopharmacol* 1993;**7**:191–5.
9. West R. Depression. London: Office of Health Economics; 1992. Report No.: 105 (abstract).
10. Lloyd K, Jenkins R. The economics of depression in primary care: Department of Health initiatives. *Br J Psychiatry* 1994;**66**:60–2.
11. Simon GE, Barber C, Birnbaum HG, Frank RG, Greenberg PE, Rose RM, *et al.* Depression and work productivity: the comparative cost of treatment versus non-treatment. *J Occup Environ Med* 2001;**43**:2–9.
12. Greenberg L, Elliott R, Lietaer G. Research on experiential psychotherapies. In: Garfield SL, Bergin AE, editors. Handbook of psychotherapy and behavior change. New York: John Wiley & Sons; 1994.
13. Priest RG, Vize C, Roberts A, Roberts M, Tylee A. Lay peoples' attitudes to treatment of depression: results of opinion poll for Defeat Depression Campaign just before its launch. *BMJ* 1996;**313**:858–9.
14. Simon GE, Von Korff M, Wagner EH, Barlow W. Patterns of antidepressant use in community practice. *Gen Hosp Psychiatry* 1993;**15**:399–408.
15. Scott J. Psychological treatments for depression. *Br J Psychiatry* 1995;**167**:289–92.
16. Bellack AS, Hersen M, Himmelhoch JH. A comparison of social-skills training, pharmacotherapy and psychotherapy for depression. *Behav Res Ther* 1983;**21**:101–7.
17. Beck AT, Rush AJ, Shaw BF, Emery G. Cognitive therapy of depression. New York: Guildford Press; 1979.
18. Neimeyer RA, Robinson LA, Berman JS, Haykal RF. Clinical outcome of group therapies for depression. *Group Analysis* 1989;**22**:73–86.
19. Klerman G, Weissman M. New applications of interpersonal psychotherapy. Washington, DC: American Psychiatric Press; 1993.
20. Netten A, Dennett J, Knight J. Unit costs of health and social care. Canterbury: Personal Social Services Research Unit, University of Kent at Canterbury; 1998.
21. Fahey T, Wessely S. Should purchasers pay for psychotherapy? *BMJ* 1993;**307**:576–7.
22. Mellor-Clark J. Psychological Therapies Research Centre, University of Leeds, Leeds (personal communication; 1998).
23. Feltham C. Time-limited counselling. London: Sage; 1997.
24. Ursano RJ, Hales RE. A review of brief individual psychotherapies. *Am J Psychiatry* 1986;**143**:1507–15.
25. Hudson-Allez G. Time-limited therapy in a general practice setting. London: Sage; 1997.
26. Jarrett RB, Rush AJ. Short-term psychotherapy of depressive disorders: current status and future directions. *Psychiatry* 1994;**57**:115–33.
27. Dryden W. Handbook of individual therapy. London: Sage; 1996.
28. Butler G, Low J. Short-term psychotherapy. In: Clarkson P, Pokorny M, editors. The handbook of psychotherapy. London: Routledge; 1994.
29. Howard KI, Kopta SM, Krause MS, Orlinsky DE. The dose–effect relationship in psychotherapy. *Am Psychol* 1986;**41**:159–64.
30. Smith ML, Glass GV. Meta-analysis of psychotherapy outcome studies. *Am Psychol* 1977;**32**:752–60.
31. Sledge WH, Moras K, Hartley D, Levine M. Effect of time-limited psychotherapy on patient dropout rates. *Am J Psychiatry* 1990;**147**:1341–7.

32. Hollon SD. System for rating psychotherapy audiotapes. Final report to NIMH. Rockville, MD: National Technical Information Service, US Department of Commerce; 1984. Contract No.: 278-81-0031.

33. McLean PD, Anderson KW. Common determinants in empirically supported psychosocial treatments for depression. In: Dobson KS, Craig KD, editors. Empirically supported therapies: best practice in professional psychology. Thousand Oaks, CA: Sage; 1998.

34. Robinson LA, Berma JS, Neimeyer RA. Psychotherapy for the treatment of depression: a comprehensive review of controlled outcome research. *Psychol Bull* 1990;**108**:30–49.

35. Gaffan EA, Tsaousis I, Kemp-Wheeler SM. Researcher allegiance and meta-analysis: the case for cognitive therapy for depression. *J Consult Clin Psychol* 1995;**63**:966–80.

36. Miller N, Magruder K, editors. Cost-effectiveness of psychotherapy. Oxford: Oxford University Press; 1999.

37. Healey A, Knapp M. Economic appraisal of psychotherapy. *Ment Health Res Rev* 1995;**2**:13–16.

38. Spitzer R, Endicott J, Robins E. Research Diagnostic Criteria for a selected group of functional disorders. 3rd ed. New York: Biometric Research; 1977.

39. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. 3rd ed. revised (DSM-IIIR). Washington, DC: American Psychiatric Association; 1987.

40. World Health Organization. The ICD-10 classification of mental and behaviour disorders. Diagnostic criteria for research. Geneva: WHO; 1993.

41. Beck AT, Ward C, Mendelson M, Mock J, Erbaugh J. An inventory for measuring depression. *Arch Gen Psychiatry* 1961;**4**:561–71.

42. Dahlstrom WG, Welsh GS, Dahlstrom LE. An MMPI handbook. Research, development and applications. Vol 1. Minneapolis, MN: University of Minnesota Press; 1975.

43. Hamilton M. A rating scale for depression. *J Neurology, Neurosurgery Psychiatry* 1960;**23**:56–62.

44. Moncrieff J, Churchill R, Drummond C, McGuire H. Development of a quality assessment instrument for trials of treatments for depression and neurosis. *Int J Methods Psychiatric Res* 2001;**10**:3.

45. Agency for Health Care Policy and Research. Depression in primary care. Treatment of major depression. Vol 2. Rockville, MD: Agency for Health Care Policy and Research, US Department of Health and Human; 1993. Report No.: 93-0550.

46. Rehm L. A self-control model of depression. *Behav Ther* 1977;**8**:787–804.

47. Nezu AM, Nezu CM, Perri MG. Problem-solving therapy for depression. New York: John Wiley & Sons; 1989.

48. Lewinsohn PM, Antonuccio DO, Steinmetz JL, Teri L. The coping with depression course: a psychoeducational intervention for unipolar depression. Eugene, OR: Castalia Publishing; 1984.

49. Sullivan HS, editor. The interpersonal theory of psychiatry. New York: WW Norton; 1953.

50. Klerman GL, DiMascio A, Weissman MM, Prusoff BA, Paykel ES. Treatment of depression by drugs and psychotherapy. *Am J Psychiatry* 1974;**131**:186–91.

51. Malan DH. A study of brief psychotherapy. London: Plenum; 1963.

52. Balint M, Orstien PO, Balint E. Focal psychotherapy. London: Tavistock; 1972.

53. Mann J. Time-limited therapy. Cambridge, MA: Havard University Press; 1973.

54. Davanloo H, editor. Basic principles and techniques in short-term dynamic psychotherapy. London: Spectrum; 1980.

55. Luborsky L. Principles of psychoanalytic psychotherapy: a manual for supportive-expressive treatment. New York: Basic Books; 1984.

56. Strupp HH, Binder JL. Psychotherapy in a new key: a guide to time-limited dynamic psychotherapy. New York: Basic Books; 1984.

57. Rogers C. On becoming a person. Boston, MA: Houghton Mifflin; 1961.

58. Dalrup RJ, Beutler LE, Engle D, Greenberg LS. Focussed expressive psychotherapy: freeing the overcontrolled patient. New York: Guilford Press; 1988.

59. Von Korff M, Katon W, Bush T, Lin EM, Simons GE, Saunders K, *et al.* Treatment costs, cost offset and cost-effectiveness of collaborative management of depression. *Psychosom Med* 1998;**60**:143–9.

60. Lave JR, Frank RG, Schulberg HC, Kamlet MS. Cost-effectiveness of treatments for major depression in primary care practice. *Arch Gen Psychiatry* 1998;**55**:645–51.

61. Prioleau L, Murdock M, Brody N. An analysis of psychotherapy versus placebo studies. *Behav Brain Sci* 1983;**6**:275–310.

62. Weishaar ME. Aaron T Beck. Thousand Oaks, CA: Sage; 1993.

63. Hennekins CH, Buring JE. Epidemiology in medicine. Boston, MA: Little Brown; 1997.

64. Posternak MA, Zimmerman M. Short-term spontaneous improvement rates in depressed outpatients. *J Nerv Ment Dis* 2000;**188**:799–804.

65. Lewinsohn PM, Biglan A, Zeiss AM. Behavioural treatment of depression. In: Davidson PO, editor. The behavioural management of anxiety, depression and pain. New York: Brunmer/ Mazel, 1976.

66. Lubin B. Adjective checklists for the measurement of depression. *Arch Gen Psychiatry* 1965;**12**:57–62.

67. Palazzoli MS, Boscolo L, Cecchin G, Prata G. Paradox and counterparadox. New York: Jason Aronson, 1978.

68. Beck AT. Cognitive therapy and the emotional disorders. New York: International University Press; 1976.

69. Ellis A. Reason and emotion in psychotherapy. New York: Lyle Stuart; 1962.

70. Barak A, La Crosse MB. Multidimensional perception of counsellor behaviour. *J Counsel Psychol* 1975;**22**:471–6.

71. Richardson A. Verbaliser–visualiser: a cognitive style dimension. *J Mental Imagery* 1977;**1**:109–26.

72. Spitzer R, Williams J, Gibbons M, First M. Structured Clinical Interview for DSM III R. Washington, DC: American Psychiatric Association; 1989.

73. Battle C, Imbers S, Hoen-Soric R, Stone A, Nash C, Frank J. Target complaints as criteria for improvement. *Am J Psychol* 1968;**20**:184–92.

74. Costa PT, McCrae RR. The NEO Personality Inventory Manual. Odessa, FL: Psychological Assessment Resources; 1985.

75. Fuchs CZ, Rehm LP. A self-control behavior therapy program for depression. *J Consult Clin Psychol* 1977;**45**:206–15.

76. Friedenheit AR. Statistical analysis and norms for an adjustment inventory [unpublished Masters Thesis]. New York: Hunter College; 1969. (Scale available from the Institute for Advanced Study in Rational Psychotherapy.)

77. Feighner JP, Robins E, Guze SB, Woodruff RA, Winokur RA, Munoz R. Diagnostic criteria for use in psychiatric research. *Arch Gen Psychiatry* 1972;**26**:57–63.

78. Shostrom EL. A test for the measurement of self-actualisation. *Educational Psychological Measurement* 1965;**24**:207–18.

79. Rosenbaum M. A schedule for assessing self-control behaviours: preliminary findings. *Behav Res Ther* 1980;**11**:109–21.

80. Nezu AM. Effects of stress from current problems: comparison to major life events. *J Consult Clin Psychol* 1986;**54**:196–202.

Appendix 1

Search strategies used for electronic databases

CCDANCTR

(#30 = BEHAVIOR-THERAPY or
#30 = BIOFEEDBACK or
#30 = CASE-MANAGEMENT or
#30 = COGNITIVE-ANALYTIC-THERAPY or
#30 = COGNITIVE-BEHAVIOR-THERAPY or
#30 = COGNITIVE-THERAPY or
#30 = COUNSELLING or
#30 = FAMILY-THERAPY or
#30 = MARITAL-THERAPY or
#30 = PSYCHOANALYTIC-THERAPY or
#30 = PSYCHOTHERAPY or
#30 = RELAXATION-THERAPY or
#30 = SOCIAL-INTERVENTION)
AND
(#45 = DEPRESS* or #45 = DYSTHYMI*)
where #30 = intervention field and #45 = diagnostic field

CCDANCTR incorporated the results of the following searches:

The search strategy, using SilverPlatter MEDLINE, was as follows:

SilverPlatter ASCII 3.0 WINN Selected Databases

#1 explode "Eating-Disorders"/all subheadings
#2 explode "Anorexia-Nervosa"/all subheadings
#3 explode "Bulimia"/all subheadings
#4 explode "Obesity"/all subheadings
#5 explode "Suicide,-Attempted"/all subheadings
#6 explode "Self-Mutilation"/all subheadings
#7 explode "Self-Injurious-Behavior"/ all subheadings
#8 explode "Affective-Disorders"/all subheadings
#9 explode "Bipolar-Disorder"/all subheadings
#10 explode "Manic-Disorder"/all subheadings
#11 explode "Neurotic-Disorders"/all subheadings
#12 explode "Depression"/all subheadings
#13 explode "Depressive-Disorder"/ all subheadings
#14 explode "Seasonal-Affective-Disorder"/ all subheadings
#15 explode "Adjustment-Disorders"/ all subheadings
#16 explode "Anxiety"/all subheadings
#17 explode "Anxiety-Disorders"/all subheadings
#18 explode "Panic"/all subheadings
#19 explode "Panic-Disorder"/all subheadings
#20 explode "Phobic-Disorders"/all subheadings
#21 explode "Stress-Disorders,-Post-Traumatic"/ all subheadings
#22 explode "Combat-Disorders"/all subheadings
#23 explode "Somatoform-Disorders"/ all subheadings
#24 explode "Hypochondriasis"/all subheadings
#25 explode "Hysteria"/all subheadings
#26 explode "Conversion-Disorder"/ all subheadings
#27 explode "Munchausen-Syndrome"/ all subheadings
#28 explode "Munchausen-Syndrome-by-Proxy"/ all subheadings
#29 explode "Neurasthenia"/all subheadings
#30 explode "Fatigue-Syndrome,-Chronic"/ all subheadings
#31 explode "Obsessive-Compulsive-Disorder"/ all subheadings
#32 explode "Psychosexual-Disorders"/ all subheadings
#33 explode "Impotence"/all subheadings
#34 explode "Frigidity"/all subheadings
#35 explode "Dysthymic-Disorder"/ all subheadings
#36 explode "Affective-Symptoms"/ all subheadings
#37 explode "Stress,-Psychological"/ all subheadings
#38 explode "Mental-Disorders"/all subheadings
#39 explode "Mood-Disorders"/all subheadings
#40 explode "Obsessive-Behavior"/all subheadings
#41 explode "Compulsive-Behavior"/ all subheadings
#42 #1 or #2 or #3 or #4 or #5 or #6 or #7 or #8 or #9 or #10 or #11 or #12 or #13 or #14 or #15 or #16 or #17 or #18 or #19 or #20 or #21
#43 #22 or #23 or #24 or #25 or #26 or #27 or #28 or #29 or #30 or #31 or #32 or #33 or #34 or #35 or #36 or #37 or #38 or #39 or #40 or #41
#44 #42 or #43 combined with the Cochrane Collaboration optimal RCT search strategy.

EMBASE

EMBASE (1980–1998) was searched using the following search strategy, and was updated quarterly.

SilverPlatter ASCII 3.0 WINN EMBASE CD (R) July 1998–November 1998

#1 CONTROLLED-STUDY
#2 CONTROLLED-STUDY in DE
#3 CLINICAL-TRIAL
#4 CLINICAL-TRIAL in DE
#5 MAJOR-CLINICAL-STUDY
#6 MAJOR-CLINICAL-STUDY in DE
#7 RANDOMIZED-CONTROLLED-TRIAL
#8 RANDOMIZED-CONTROLLED-TRIAL in DE
#9 DOUBLE-BLIND-PROCEDURE
#10 DOUBLE-BLIND-PROCEDURE in DE
#11 CLINICAL-ARTICLE
#12 CLINICAL-ARTICLE in DE
#13 RANDOM*
#14 TRIAL*
#15 COMPAR*
#16 CONTROL*
#17 STUDY
#18 FOLLOW*
#19 UP
#20 FOLLOW* and UP
#21 CLINIC*
#22 BLIND*
#23 PLACEBO*
#24 DOUBL*
#25 #2 or #4 or #6 or #8 or #10 or #12 or #13 or #14 or #15 or #16 or #17 or #18 or #19 or #20 or #21 or #22 or #23 or #24
#26 ANIMAL
#27 NON-HUMAN
#28 ANIMAL or NON-HUMAN
#29 HUMAN
#30 #28 and #29
#31 #28 not #30
#32 #25 not #31
#33 explode "neurosis"/all subheadings
#34 explode "mania"/all subheadings
#35 explode "manic-depressive-psychosis"/all subheadings
#36 explode "anorexia"/all subheadings
#37 explode "anorexia-nervosa"/all subheadings
#38 explode "bulimia"/all subheadings
#39 explode "obesity"/all subheadings
#40 explode "suicidal-behavior"/all subheadings
#41 explode "chronic-fatigue-syndrome"/all subheadings
#42 explode "psychosexual-disorder"/all subheadings
#43 explode "frigidity"/all subheadings
#44 explode "automutilation"/all subheadings
#45 explode "anxiety"/all subheadings
#46 explode "depression"/all subheadings
#47 #33 or #34 or #35 or #36 or #37 or #38 or #39 or #40 or #41 or #42 or #43 or #44 or #45 or #46
#48 #32 and #47

PsycLIT

PsycLIT (1974–1995) was searched from 1996 to 1997, and the following search was carried out by CCDAN and was updated quarterly.

SilverPlatter ASCII 3.0 WINN PsycLIT Journal Articles 1991–December 1997

#1 RANDOM*
#2 SINGL*
#3 DOUBL*
#4 TRIPL*
#5 TREBL*
#6 BLIND*
#7 MASK*
#8 (SINGL* or DOUBL* or TRIPL* or TREBL*) near (BLIND* or MASK*)
#9 CROSSOVER
#10 CROSS-OVER
#11 VERSUS
#12 VS
#13 #1 or #8 or #9 or #10 or #11 or #12

LILACS

LILACS (1982–1996) was searched using the following search strategy and was updated quarterly.

#1 RANDOM$
#2 ALEATORI$ or CASUAL or ACASO or AZAR
#3 ((DUPLO or DOBLE or SIMPLE or TRIPLO or TRIPLE) and (CEGO or CIEGO))
#4 ((DOUBL$ or SINGL$ or TRIPL$ or TREBL$) and (BLIND$ or MASK$)
#5 SINGLE-MASKED STUDY/
#6 DOUBLE-MASKED STUDY/
#7 PROPHYLATIC CONTROLLED TRIALS/
#8 PLACEBO$ and CONTROL$
#9 CLINICAL$ and TRIAL$
#10 #1 or #2 or #3 or #4 or #5 or #6 or #7 or #8 or #9

PSYNDEX

PSYNDEX (1977–1995) was searched by the Cochrane Schizophrenia Group using the following search strategy.

#1 RANDOM* and (ALLOCAT* or ASSIGN*)
#2 RANDOMI*
#3 (DOUBL* or SINGL* or TRIPL* or TREBL*) near (BLIND* or MASK*)
#4 DOPPELBLIND*
#5 PLA?EBO* and ((EITHER or ENTWEDER) or (TREAT* or BEHAND* or UNTERSUCH*))
#6 PLA?EBO* near ((VS or VERSUS) or VERUM)
#7 [ZUFA?LL* and (EXPERIMENT* or EVALU* or EFFE?T*) and TREAT*]
#8 ZUGEWIESEN and KONTROLLGRUPPE*
#9 #1 or #2 or #3 or #4 or #5 or #6 or #7 or #8

CINAHL

This database (1982–1997) was searched using the following strategy and was updated quarterly.

SilverPlatter ASCII 3.0 WINN CINAHL (R) Database 1982–September 1997

#1 explode "Depression"/all topical subheadings/all age subheadings
#2 explode "Eating-Disorders"/all topical subheadings/all age subheadings
#3 explode "Anorexia-Nervosa"/all topical subheadings/all age subheadings
#4 explode "Anorexia"/all topical subheadings/all age subheadings
#5 explode "Bulimia"/all topical subheadings/all age subheadings
#6 explode "Suicide,-Attempted"/all topical subheadings/all age subheadings
#7 explode "Injuries,-Self-Inflicted"/all topical subheadings/all age subheadings
#8 explode "Affective-Disorders"/all topical subheadings/all age subheadings
#9 explode "Bipolar-Disorder"/all topical subheadings/all age subheadings
#10 explode "Affective-Disorders,-Psychotic"/all topical subheadings/all age subheadings
#11 explode "Neurotic-Disorders"/all topical subheadings/all age subheadings
#12 explode "Depression,-Postpartum"/all topical subheadings/all age subheadings
#13 explode "Depression,-Reactive"/all topical subheadings/all age subheadings
#14 explode "Seasonal-Affective-Disorder"/all topical subheadings/all age subheadings
#15 explode "Adjustment-Disorders"/all topical subheadings/all age subheadings
#16 explode "Anxiety"/all topical subheadings/all age subheadings
#17 explode "Anxiety-Disorders"/all topical subheadings/all age subheadings
#18 explode "Dental-Anxiety"/all topical subheadings/all age subheadings
#19 explode "Panic-Disorder"/all topical subheadings/all age subheadings
#20 explode "Phobic-Disorders"/all topical subheadings/all age subheadings
#21 explode "Stress-Disorders,-Post-Traumatic"/all topical subheadings/all age subheadings
#22 explode "Somatoform-Disorders"/all topical subheadings/all age subheadings
#23 explode "Hypochondriasis"/all topical subheadings/all age subheadings
#24 explode "Hysteria"/all topical subheadings/all age subheadings
#25 explode "Munchausen-Syndrome"/all topical subheadings/all age subheadings
#26 explode "Munchausen-Syndrome-By-Proxy"/all topical subheadings/all age subheadings
#27 explode "Fatigue-Syndrome,-Chronic"/all topical subheadings/all age subheadings
#28 explode "Obsessive-Compulsive-Disorder"/all topical subheadings/all age subheadings
#29 explode "Psychosexual-Disorders"/all topical subheadings/all age subheadings
#30 explode "Impotence"/all topical subheadings/all age subheadings
#31 explode "Frigidity"/all topical subheadings/all age subheadings
#32 #1 or #2 or #3 or #4 or #5 or #6 or #7 or #7 or #8 or #9 or #10 or #11 or #12 or #13 or #14 or #15 or #16 or #17 or #18 or #19 or #20 or #21 or #22 or #23 or #24 or #25 or #26 or #27 or #28 or #29 or #30 or #31

Cochrane Library

The search strategy for the Cochrane Library has been split into two sections (shown below) and was updated quarterly.

CCDAN1

#1 (EATING and DISORDER*)
#2 (ANOREXIA and NERVOSA)
#3 BULIMI*
#4 OBESITY
#5 (ATTEMPTED and SUICIDE)
#6 (SELF and MUTILAT*)
#7 (SELF and INJUR*)
#8 (AFFECTIVE and DISORDER*)
#9 (BIPOLAR and DISORDER*)
#10 (MANIC and DISORDER*)
#11 MANIA
#12 (NEUROTIC and DISORDER*)
#13 DEPRESSION
#14 DEPRESSED
#15 DEPRESSIVE
#16 (ADJUSTMENT and DISORDER*)

#17 ANXIETY
#18 (ANXIETY and DISORDER*)
#19 PANIC*
#20 (PHOBIC and DISORDER*)
#21 PHOBIA
#22 (STRESS and DISORDER*)
#23 ((((((((((((((((((((#1 or #2) or #3) or #4) or #5) or #6) or #7) or #8) or #9) or #10) or #11) or #12) or #13) or #14) or #15) or #16) or #17) or #18) or #19) or #20) or #21) or #22)

CCDAN2
#1 (COMBAT and DISORDER*)
#2 (SOMATOFORM and DISORDER*)
#3 PSYCHOSOMAT*
#4 HYPOCHONDRIASI*
#5 HYSTERI*
#6 (CONVERSION and DISORDER*)
#7 MUNCHAUSEN*
#8 NEURASTHENI*
#9 (CHRONIC and FATIGUE)
#10 CFS
#11 OCD
#12 OBSESS*
#13 COMPULS*
#14 (PSYCHOSEXUAL and DISORDER*)
#15 IMPOTEN*
#16 FRIGID*
#17 DYSTHYMI*
#18 STRESS*
#19 (AFFECTIVE and SYMPTOMS)
#20 ((((((((((((((((((#1 or #2) or #3) or #4) or #5) or #6) or #7) or #8) or #9) or #10) or #11) or #12) or #13) or #14) or #15) or #16) or #17) or #18) or #19)

Appendix 2

Data extraction spreadsheet for the studies included in the review

Study number*	Design	Diagnosis	Intervention A	Intervention B	Intervention C	Intervention D	Intervention E	Individual/ group	Country	Intervention site	Method of recruitment	Mean age	Therapist qualifications	Specialist training
1	RCT	MDD	CBT (eight sessions; *n* = unspecified)	IPT/PDT (eight sessions; *n* = unspecified)	CBT (16 sessions; *n* = unspecified)	IPT/PDT (16 sessions; *n* = unspecified)	–	Individual	UK	Three NHS outpatient departments	Referral letter	38.7	Clinical psychologists	Weekly for 6 months
2	RCT	Clinical depression	BT (*n* = 10)	Waiting list (*n* = 10)	–	–	–	Group	USA	University psychology department	Referrals and media advertisements	35.8	Clinical psychology students	NR
3	RCT	Negative mood state	Positive connotation (*n* = 10)	Negative connotation (*n* = 10)	No interview (*n* = 10)	–	–	Individual	USA	University psychology department	Student volunteers	20.0	Counselling psychology students	12 hrs of training
4	RCT	MDD	BT + placebo (*n* = 33)	PDT + placebo (*n* = 31)	(BT + antidepressant)	(Anti-depressant)	–	Individual	USA	Psychiatry department	Referrals and media advertisements	35.7 (females only)	PhD psychologists	NR
5	RCT	MDD	CT (*n* = 27)	Focused expressive (*n* = 28)	(ST by telephone)	–	–	Group	USA	University psychology department	Media advertisements and referrals	46.8	Psychologists/ students	NR
6	RCT	MDD/ minor depressive disorder	Individual BT (*n* = 13)	Group BT (*n* = 25)	Waiting list (*n* = 11)	(Minimum contact by telephone)	–	Individual/ group	USA	University psychology department	Media advertisements	36.5	Advanced clinical psychology students	Trained in pilot study
7	CCT	Dysphoria	Overt/ covert reward (*n* = 42)	Overt punishment (*n* = 19)	Covert punishment (*n* = 19)	Social influence (*n* = 19)	Waiting list (*n* = 21)	Group	USA	University psychology department	Student volunteers	NR	Therapist	NR
8	RCT	Unipolar depression	BT (*n* = 8)	CT (*n* = 8)	Waiting list (*n* = 10)	–	–	Group	USA	Mental health clinic	Community agency referrals	38.0	Therapist	NR
9	RCT	Moderate depression	Self-control (*n* = 19)	Reframing (*n* = 19)	Waiting list (*n* = 19)	–	–	Individual	USA	University psychology department	Student volunteers	NR	PhD students	6 hrs of training
10	RCT	Mild/ moderate depression	PDT (*n* = unspecified)	ST (*n* = unspecified)	PDT + placebo	ST + placebo	(Therapy + medication)	Group	USA	Hospital outpatients	Outpatient attenders	34.4 (females only)	Psychiatrists	NR
11	RCT	MDD	CBT (*n* = 32)	PDT/IPT (*n* = 24)	(CBT + antidepressant)	–	–	Group	USA	Psychiatry department	Media advertisements	43.8	Psychiatrist/ psychologist	NR
12	RCT	MDD	IPT (*n* = 25)	ST (*n* = 23)	(Anti-depressant)	(IPT + antidepressant)	–	Individual	USA	Hospital/ mental health clinic	Outpatient attenders	55% over 30.0	Psychiatrist	NR

continued

Study number*	Design	Diagnosis	Intervention A	Intervention B	Intervention C	Intervention D	Intervention E	Individual/ group	Country	Intervention site	Method of recruitment	Mean age	Therapist qualifications	Specialist training
13	RCT	MDD	CBT (*n* = 62)	IPT (*n* = 63)	Clinical management/ placebo (*n* = 62)	(Anti-depressant)	–	Individual	USA	Three university sites	Self + referred patients	35.0	Psychiatrists/ psychologists	2 yrs of training
14	RCT	Negative mood state	Paradoxical (*n* = 10)	Non-paradoxical (*n* = 9)	Combination 1 (*n* = 10)	Combination 2 (*n* = 10)	No treatment (*n* = 10)	Individual	USA	University psychology department	Student volunteers	22.0	Counselling psychologists	10 hrs of training
15	RCT	Clinical depression	CT (*n* = 13)	BT (*n* = 13)	Non-directive (*n* = 9)	–	–	Group	USA	University psychology department	Media advertisements	38.0	Clinical psychology students	4 hrs of training
16	RCT	Depression	BT (*n* = 12)	Non-specific (*n* = 12)	Waiting list (*n* = 12)	–	–	Group	USA	University psychology centre	Media advertisements	28.8 (females only)	Psychologists	Some training sessions given
17	RCT	MDD/ minor depressive disorder	CBT (*n* = 36)	PDT (*n* = 30)	–	–	–	Individual	USA	Older adult centre	Referrals and media advertisements	62.0	PhD psychologists	No
18	RCT	Self-rated depression	CT (*n* = 8)	BT (*n* = 8)	–	–	–	Group	USA	University psychology department	Media advertisements	19.0–65.0	Therapist	NR
19	RCT	Self-rated depression	Positive imagery (*n* = 10)	Positive imagery/self-elaboration (*n* = 10)	No treatment (*n* = 10)	–	–	Individual	USA	University psychology department	Student volunteers	NR	Experimenter	NR
20	RCT	MDD	Process experiential (*n* = 17)	Client-centred (*n* = 17)	–	–	–	Individual	USA	University psycho-therapy research centre	Media advertisements	39.6	Range of experience/ qualifications	24 wks of training
21	RCT	Mild/ moderate depression	Assertion (*n* = 16)	Waiting list (*n* = 12)	–	–	–	Group	USA	University psychology department	University referrals/ advertisements	21.3	Counselling psychology students	Yes, but no details given
22	RCT	Unipolar depression	CT (*n* = 13)	IPT (*n* = 14)	(Non-concurrent control)	–	–	Group	USA	University psychology department	Psychology department attenders	23.1	Counselling psychology students	In supervision

continued

Study number*	Design	Diagnosis	Intervention A	Intervention B	Intervention C	Intervention D	Intervention E	Individual/ group	Country	Intervention site	Method of recruitment	Mean age	Therapist qualifications	Specialist training
23	RCT	MDD	CT (*n* = 50)	(Behavioural activation/ automatic thoughts)	BT (*n* = 57)	–	–	Individual	USA	University psychology department	Referred (health maintenance organisation) and media advertisements	37.9	Experienced cognitive therapists	1 yr of training
24	CCT	Self-rated depression	Neutral imagery (*n* = unspecified)	Positive imagery (*n* = unspecified)	Combination (*n* = unspecified)	Waiting list (*n* = 8)	–	Individual	USA	University psychology department	Student volunteers	NR	Experimenter	NR
25	RCT	MDD	CBT (*n* = 77)	Treatment as usual (*n* = 76)	–	–	–	Individual	USA	Primary care clinic	Primary care attenders	46.0	Psychologist/ physician	Yes, series of meetings
26	CCT	MAD	Self-control (*n* = 16)	Self-monitoring/ evaluation (*n* = 12)	Principles only (*n* = 15)	PDT (*n* = 6)	–	Group	USA	University psychology department	Media advertisements	37.9 (females only)	Advanced clinical psychology students	20 hrs of training
27	RCT	Self-rated depression	CT (*n* = 12)	Assertion (*n* = 10)	PDT (*n* = 11)	–	–	Group	USA	University psychology department	Media advertisements	35.1 (females only)	Interns	NR
28	RCT	Unipolar depression	BT (*n* = 44)	PDT (*n* = 51)	Relaxation (*n* = 48)	(Anti-depressant)	–	Individual	Canada	Psychiatric hospital outpatients	Media advertisements	39.2	Psychologists/ psychiatrists	Yes, but no details given
29	RCT	Depression	CBT (*n* = 12)	CT (*n* = 13)	BT (*n* = 13)	Client-centred (*n* = 12)	–	Individual	USA	University counselling centre	Counselling centre attenders	23.0	Interns	8 hrs of training
30	RCT	Neurotic depression	BT (*n* = 27)	Client-centred (*n* = 30)	Waiting list (*n* = 14)	–	–	Group	Canada	Mental health clinic	Mental health clinic	35.4	Social workers	NR
31	RCT	MAD	CBT (*n* = 11)	Relaxation (*n* = 14)	(Anti-depressant)	–	–	Individual	USA	University medical school	Media/ university advertisements	39.4	Psychologists/ psychiatrist	Trained over several months
32	RCT	MDD	Combined CT (*n* = 72)	IPT (*n* = 39)	–	–	–	Group	USA	University psychology clinic	Media advertisements	43.4	PhD psychologists/ students	No

continued

Study number*	Design	Diagnosis	Intervention A	Intervention B	Intervention C	Intervention D	Intervention E	Individual/ group	Country	Intervention site	Method of recruitment	Mean age	Therapist qualifications	Specialist training
33	RCT	Unipolar depression	Problem-solving (*n* = 12)	Problem-focused (*n* = 11)	Waiting list (*n* = 9)	–	–	Group	USA	University mental health clinic	Media advertisements	41.7	Advanced clinical psychology students	Yes, but no details given
34	RCT	MDD	Problem-solving (*n* = 15)	Abbreviated problem-solving (*n* = 14)	Waiting list (*n* = 13)	–	–	Group	USA	University psychology department	Media advertisements	45.8	Clinical psychology graduate students	Yes, but no details given
35	RCT	Mild/ moderate depression	CT (*n* = 44)	No treatment (*n* = 45)	–	–	–	Individual	USA	University psychology department	Student volunteers	22.2	Graduate counselling students	12-hr training workshop
36	CCT	Moderate depression	BT (*n* = 12)	Client-centred (*n* = 11)	–	–	–	Group	USA	University psychology department	Media advertisements/ communication agencies	21.0–56.0 (females only)	Counsellor	NR
37	RCT	Depression	Self-control (*n* = 14)	Assertion (*n* = 13)	–	–	–	Group	USA	University psychology department	Media advertisements	21.0–60.0 (females only)	PhD clinical psychology students	No
38	RCT	MAD	Self-monitoring (*n* = 12)	Self-monitoring/ evaluation (*n* = 11)	Self-monitoring/ reinforcement (*n* = 12)	Self-control (*n* = 12)	Waiting list (*n* = 16)	Group	USA	University psychology department	Media advertisements	39.2 (females only)	Advanced clinical psychology students	8 hrs of training
39	RCT	MAD	CBT (n ≥ 35)	CT (n ≥ 35)	BT (n ≥ 34)	–	–	Group	USA	University psychology department	Media advertisements	38.6 (females only)	PhD-level psychologists	In supervision
40	RCT	MDD	Individual CT (*n* = 27)	Group CT (*n* = 23)	Waiting list (*n* = 23)	–	–	Individual/ group	UK	Primary care clinic	Health centre attenders	33.0	Social worker	No
41	CCT	Depressive syndrome	Individual CT (*n* = 9)	Group CT (*n* = 28)	(CT + antidepressant)	–	–	Individual/ group	USA	University School of Medicine	Outpatient attenders?	39.2	Mixed: psychiatrists to interns	NR
42	RCT	Clinical depression	Assertion (*n* = 16)	PDT (*n* = 16)	–	–	–	Group	USA	University medical centre outpatients department	Psychiatric outpatients department attenders	NR	Advanced clinical psychology students	No

continued

Study number*	Design	Diagnosis	Intervention A	Intervention B	Intervention C	Intervention D	Intervention E	Individual/ group	Country	Intervention site	Method of recruitment	Mean age	Therapist qualifications	Specialist training
43	RCT	Clinical depression	Individual CBT (*n* = 12)	Small group CBT (*n* = 11)	Large group CBT (*n* = 11)	(Bibliotherapy)	Waiting list (*n* = 10)	Individual/ group	USA	University psychology department	Media feature article	42.0	Para-professional therapists	Yes, but no details given
44	RCT	MDD	IPT (*n* = 93)	Treatment as usual (*n* = 92)	(Anti-depressant)	–	–	Individual	USA	Primary care clinics	Primary care attenders	37.9	Psychiatrists/ psychologists	16 hrs of training
45	RCT	MDD	CBT (*n* = 30)	ST (*n* = 30)	Treatment as usual (*n* = 30)	(Anti-depressant)	–	Individual	UK	Primary care practices	Primary care attenders	31.8	Clinical psychologist/ social worker	No
46	RCT	MDD	CT (*n* = 24)	Treatment as usual (*n* = 24)	–	–	–	Individual	UK	Primary care practices	Primary care attenders	41.0	Cognitive therapist/ GP	No
47	RCT	MDD	Individual CT (*n* = 13)	Group CT (*n* = 23)	–	–	–	Individual/ group	UK	Primary care practices?	Occupational health/GP referrals	36.0	Psychologist	No
48	RCT	MDD/ minor depressive disorder	CBT (*n* = 12)	Computer CBT (*n* = 12)	Waiting list (*n* = 12)	–	–	Individual	USA	Institute of technology	Media advertisements	28.2	Advanced clinical psychology students	Yes, but no details given
49	RCT	MDD	CBT (eight sessions; *n* = 35)	IPT/PDT (eight sessions; *n* = 35)	CBT (16 sessions; *n* = 40)	IPT/PDT (16 sessions; *n* = 40)	–	Individual	UK	University psychology clinic	GP/ occupational health referrals	40.5	Research clinical psychologists	Clinical training
50	RCT	Clinical depression	CT (*n* = 8)	BT (*n* = 8)	Client-centred (*n* = 8)	Waiting list (*n* = 8)	–	Group	Canada	University psychology department	Student health service/self-referrals	18.0–26.0	Clinical psychology students	No
51	CCT	Self-rated depression	BT (*n* = 11)	Waiting list (*n* = 11)	–	–	–	Individual	USA	University psychology department	Student volunteers	NR	PhD clinical psychology student	NR
52	CCT	Depression	BT expectancy (*n* = 7)	ST expectancy (*n* = 7)	BT (*n* = 7)	ST (*n* = 7)	–	Individual	USA	University psychology department	Student volunteers	NR	PhD clinical psychology student	NR
53	RCT	Mild/ moderate depression	CBT (*n* = 7)	CT (*n* = 7)	BT (*n* = 7)	Waiting list (*n* = 7)	–	Individual	USA	University psychology department	Campus advertisements	22.4	Graduate student therapist	NR
54	RCT	MDD	CT (*n* = 24)	Treatment as usual (*n* = 20)	–	–	–	Individual	UK	Primary care practices	Primary care attenders	37.5	CT trained clinical psychologists	No

continued

Study number*	Design	Diagnosis	Intervention A	Intervention B	Intervention C	Intervention D	Intervention E	Individual/ group	Country	Inter-vention site	Method of recruitment	Mean age	Therapist qualifi-cations	Special-ist training
55	RCT	MDD/ minor depressive disorder	Individual BT (*n* = 19)	Group BT (*n* = 47)	–	–	–	Individual/ group	UK	University psychology clinic	Psychology clinic attenders	34.7	Advanced clinical psychology students	3 months of training
56	CCT	Depression	BT (*n* = 14)	Client-centred (*n* = 14)	(Fitness program)	(Self-monitoring)	–	Individual	USA	University psychology department	Campus advertisements	24.5	Advanced clinical psychology student	NR
57	RCT	Mild depression	Gestalt (*n* = unspecified)	Attention-placebo (*n* = unspecified)	(No dialogue)	(Neutral dialogue)	–	Group	USA	University psychology departments	Campus/ community advertisements	24.6	PhD clinical psychology student	NR
58	RCT	MDD	CBT (*n* = 31)	Waiting list (*n* = 29)	–	–	–	Group	Australia	University psychology department	Media advertisements	39.0	Advanced clinical psychology student	In super-vision
59	RCT	Mild depression	Individual CT (*n* = 9)	Group CT (*n* = 9)	–	–	–	Individual/ group	USA	University psychology department	Local media advertisements	NR	Graduate psychology students	Yes, but no details given
60	RCT	Depression	CT (*n* = 8)	BT (*n* = 8)	Waiting list (*n* = 9)	–	–	Individual	Australia	University psychology clinic	Media advertisements	39.5	Two principal therapists	NR
61	RCT	Moderate/ severe depression	CT (*n* = unspecified)	IPT/BT (*n* = unspecified)	BT (*n* = unspecified)	–	–	Individual	USA	University psychology department	Campus and media advertisements	33.9	Graduate counselling psychology students	Yes, but no details given
62	RCT	Depression	CT ($n \geq 10$)	(Partial CT)	Compre-hensive distance ($n \geq 11$)	–	–	Group	USA	University psychology department	Local media advertisements	41.3 (females only)	Cognitive therapist/ graduate student	No
63	RCT	Depression	Individual CT (*n* = 14)	Group CT (*n* = 16)	–	–	–	Individual/ group	USA	University psychology department	Local media advertisements	43.3	Cognitive therapist	No

continued

Study number*	Monitoring of therapy	Author/ researcher	Incentives	Dichotomous outcome used	Post-treatment assessment	Post-treatment recovery reported	Final follow-up	Final follow-up recovery reported	Comments on follow-up
1	Audiotapes in supervision				8 or 16 weeks		1 year		
2	NR	Author 1 of two therapists			4 weeks		7 months		
3	Interview scripts used		Class credit		2 weeks		6 weeks		
4	NR			BDI or HRSD cut-off of 10	12 weeks	Yes	6 months		
5	Videotapes rated by expert		US$100 deposit	HRSD cut-off of 10	20 weeks	Yes	8 months	Yes	
6	Supervision discussion		Registration fee of $150	RDC for depression	8 weeks		6 months	Yes	
7	NR	Author 1 of therapists	Course credit		4 weeks		2 years (unplanned)		
8	Tapes rated by observer	Author			5 weeks		10 weeks		
9	Tapes assessed by raters				2 weeks		4 weeks		
10	NR				19 weeks		Not relevant		
11	Yes, but no details			BDI remitted/mild/severe		14 weeks	Yes	1 year	
12	NR			RGSD failure of ≥ 9	16 weeks	Yes	16 months		
13	Tapes assessed by raters			HRSD cut-off of ≤ 6	16 weeks	Yes	18 months	Yes	
14	NR		Class credit			4 weeks		8 weeks	
15	Tapes assessed by raters					4 weeks		10 weeks	
16	Supervision discussion		US$10 deposit	BDI cut-off of ≤ 10	7 weeks	Treatment groups	14 weeks		
17	Tapes assessed by raters				RDC remitted/ improved/not	16–20 weeks	Yes	1 year	
18	NR				6 weeks?		11 weeks		
19	NR	Second author		Course credit		3 weeks		None	
20	Tapes assessed by raters			BDI cut-off of ≤ 10	20 weeks	Yes	6 months	Yes	
21	NR			BDI cut-off of < 13		4 weeks	Partial	12 weeks	
22	Tapes assessed by raters					8 weeks		12–14 weeks	

continued

Study number*	Monitoring of therapy	Author/ researcher	Incentives	Dichotomous outcome used	Post-treatment assessment	Post-treatment recovery reported	Final follow-up	Final follow-up recovery reported	Comments on follow-up
23	Tapes assessed by author			BDI cut-off of ≤ 8	At termination	Yes	2 years	Yes	
24	NR	First author	US$5 to complete	4 weeks		6 months			
25	Tapes assessed by raters			50% improvement on SCL-20	7 weeks		7 months	Midpoint	
26	Tapes assessed by raters			BDI cut-off of ≤ 15	12 weeks	Yes	3 months		
27	NR	Author 1 of two therapists			6 weeks		14–16 weeks		
28	Tapes assessed by raters			BDI normal/mild/severe	10–11 weeks	Yes	27 months	At 3-month follow-up	
29	Tapes monitored			BDI cut-off of < 10	10 weeks	Yes	18 weeks		
30	NR	Author 1 of four therapists			3 weeks		6 weeks		
31	Tapes reviewed by supervisors			BDI remitted/mild/severe	16 weeks	Yes	None		
32	Videotapes in supervision		Nominal fee		10 weeks?		None		
33	Supervision discussion			BDI cut-off of < 10	8 weeks	Yes	8 months	No change	
34	Supervision discussion			BDI cut-off of < 10	10 weeks	Yes	8 months	No change	
35	One-way mirror by raters		Course credit	BDI cut-off of 10	7 weeks	Yes	11 weeks	Yes	
36	Tapes assessed by raters				12 weeks		None		
37	Supervision discussion		US$10 deposit	BDI cut-off of < 11	7 weeks	Yes	12–14 weeks		
38	Supervision discussion				10 weeks		8 months		
39	Tapes assessed by raters			SADS-RDC improvement	7 weeks	Yes	8 months		
40	No	Second author		BDI improvement/no change	12 weeks?	Yes	1 year		
41	Supervision discussion			BDI remitted/improved/not	At termination	Yes	Partial at 2 months		
42	NR				5 weeks		10 weeks		
43	Tapes reviewed in meetings		US$25 deposit		10 weeks		18 weeks		
44	One-way mirror supervision			HRSD cut-off of 7	16 weeks		8 months	Yes	
45	NR			HRSD cutoff of < 7	16 weeks	Yes	None		

continued

Study number*	Monitoring of therapy	Author/ researcher	Incentives	Dichotomous outcome used	Post-treatment assessment	Post-treatment recovery reported	Final follow-up	Final follow-up recovery reported	Comments on follow-up
46	Tapes assessed by rater			BDI or HRSD	7 weeks	Yes	1 year		
47	No	First author		BDI based on Reliable Change Index	12 or 16 weeks	Yes	6 months		
48	NR	One of the authors		BDI cut-off of ≤ 9	6 weeks	Yes	2 months	Yes	
49	Tapes assessed by raters			BDI cut-off of ≤ 8	8 or 16 weeks	Yes	1 year	Yes	
50	Tapes assessed by rater	Author		BDI cut-off of < 10	4 weeks	Yes	8 weeks		
51	NR	First author			4 weeks		Treatment group		
52	NR	First author		MMPI of < 70	5 weeks	Yes	Treatment group		
53	NR	First author			4 weeks		9 weeks		
54	Tapes assessed by raters			BDI cut-off of < 14	At termination	Yes	7 months	Yes	
55	Supervision discussion			SADS	8–12 weeks	Yes	6 months	Yes	
56	NR	First author		LDACL	4 weeks		None		
57	Tapes assessed by raters	First author	Payment/credit		4 weeks		12 weeks		
58	Tapes assessed by one author	First author	US$30 deposit	BDI cut-off of < 10	9 weeks	Yes	20 weeks	Yes (not control)	
59	Supervision discussion			BDI cut-off of ≤ 7	6 weeks	Yes	None		
60	No				10 weeks		7 months		
61	Supervision discussion				4 weeks		4 months		
62	Tapes assessed by raters	First author		BDI cut-off of ≤ 15	12 weeks	Yes	5 months	Yes	
63	NR	First author		BDI cut-off of ≤ 15	12 weeks	Yes	5 months	Yes	

continued

Study number*	Author response	Assessors	Blinded	Inter-rater reliability	Primary aim of trial	Models	Model weight	Adverse effects	Graphs	Subgroups
1	Yes, no new data	Trained assessors	Independent	No	Comparison of models	CBT/PDT–IPT	Equal	NR	Yes	No
2	No	NR			Evaluation of model	BT		NR	No	No
3	No	Experimenter	No	No	Comparison of techniques		2	No	No	
4	No	Psychiatrist for HRSD	Yes	No	Comparison of models/ drugs/both	CBT/PDT	PDT = control	11 failures?	Yes	No
5	Yes, new data given	Independent psychatric evaluation	NR	No	Predictors of positive outcome			NR	No	Yes
6	No	Advanced students trained in diag- nostic interviewing	Independent	Yes	Comparison of therapy formats	Individual/group		Seven withdrawals	Yes	Yes
7	Yes, author deceased	NR			Comparison of techniques			NR	No	No
8	No	Author?	No		Comparison of models	CT/BT	Equal	NR	No	No
9	Yes, no new data	Author?	No		Comparison of techniques			NR	No	No
10	No	Authors	No		Comparison of models/ drugs/both	PDT/ST	ST = control	NR	Yes	Yes
11	No	Independent assessor	Yes	No	Comparison of models/ drug + CBT	CBT/PDT–IPT	PDT/IPT = control	One	Yes	Yes
12	Yes, no data available	Clinical evaluator	Yes	Yes	Comparison of models/ drug/both	IPT/ST	ST = control	None	Yes	Yes
13	No	Clinical evaluators	Yes	Yes	Comparison of models/ drug/clinical management	CBT/IPT/ST	ST = control	Not in therapy arm	Yes	Yes
14	No	NR			Evaluation of technique			NR	No	No
15	No	NR			Comparison of models	CT/BT/ST	ST = control	NR	No	No
16	Yes, new data given	For group inter- action measure	Co-raters only	No	Evaluation of model	BT/ST	ST = control	Possibly	Yes	Yes
17	Yes, new data given	Independent clinicians	Yes	Yes	Comparison of models	CBT/PDT	Equal	NR	No	No
18	No	NR			Comparison of models	CT/BT	Equal	NR	Yes	No

continued

Study number*	Author response	Assessors	Blinded	Inter-rater reliability	Primary aim of trial	Models	Model weight	Adverse effects	Graphs	Subgroups
19	No	NR			Evaluation of technique			NR	No	No
20	Not contacted	NR			Process study			NR	No	No
21	Yes, no new data	NR			Evaluation of technique			NR	No	Yes
22	No	NR			Comparison of models	CT/IPT	Equal?	NR	No	No
23	No	Clinical evaluators	Yes	No	Component analysis			NR	No	Yes
24	No	First author	No	No	Evaluation of technique			NR	No	No
25	Yes, new data given	Telephone interviewers	Yes	Yes	Evaluation of model	CBT		None	Yes	Yes
26	Yes, no new data	Clinical interviewers	Yes	Yes	Component analysis			NR	No	No
27	No contact address	Assertiveness measures, psychology graduates	Partial?	No	Comparison of models	CT/BT/ST	ST = control	Five	Yes	No
28	Yes, no data available	Project staff	Independent	No	Comparison of models/ drug/placebo	BT/PDT/ST	ST = control	Not in therapy arm	Yes	Yes
29	No	NR, mailed follow-up			Component analysis			NR	No	Yes
30	No	NR			Evaluation of model	BT/ST	ST = control	NR	Yes	No
31	No	Research coordinator	Independent	Yes	Comparison of model/ drug/placebo	CBT/ST	ST = control	One	Yes	Yes
32	Yes, no data available	Non-associated clinician	Independent	No	Predictors of positive outcome			NR	No	No
33	No	NR			Evaluation of model	CBT/ST	ST = control	NR	No	No
34	No	Raters	Yes	Yes	Component analysis			NR	Yes	No
35	Yes, no new data	Authors?	No		Evaluation of model	CT		No	No	No
36	Yes, no new data	NR			Comparison of models	BT/ST	Equal	Two cases of deliberate self-harm	No	Yes
37	Yes, no new data	Two measures rated by students	No	Yes	Component analysis			NR	Yes	No

continued

Study number*	Author response	Assessors	Blinded	Inter-rater reliability	Primary aim of trial	Models	Model weight	Adverse effects	Graphs	Subgroups
38	Yes, no new data	Interviewer and clinician	Yes	Yes	Component analysis			NR	No	No
39	Yes, no new data	MA clinicians	Yes	Yes	Component analysis			NR	No	Yes
40	Yes, new data given	Psychiatrist and second author	Partial	No	Comparison of therapy formats	Individual/group		NR	No	No
41	Yes, no data available	Independent evaluator	Independent	No	Comparison of therapy formats/drug	Individual/group		NR	Yes	No
42	No	NR			Evaluation of model	BT/PDT	PDT = control	NR	No	No
43	No	Staff (not own therapist)	No	No	Comparison of therapy format	Individual/group		NR	Yes	No
44	Yes, new data given	Clinical evaluators	Yes	Yes	Evaluation of model/drug			Eight withdrawals	Yes	Yes
45	Yes, new data given	Independently trained raters	Yes	No	Comparison of models/ drug	CBT/counselling	Equal	NR	No	Yes
46	Yes, new data given	Independent assessor	Yes	No	Evaluation of model	CT		No	No	No
47	Yes, no new data	Therapist	No	No	Comparison of therapy format	Individual/group		NR	Yes	No
48	Yes, no new data	Independent interviewer	Yes	No	Evaluation of model	CBT		NR	Yes	No
49	Yes, new data given	Research interviewer	No		Comparison of models	CBT/PDT–IPT	Equal	Possibly	Yes	Yes
50	No	Independent assessor	Yes	No	Comparison of models	CT/BT/ST	Equal?	One	Yes	No
51	Yes, author deceased	NR			Evaluation of model	BT		NR	No	No
52	Yes, author deceased	NR			Comparison of models	BT/ST	ST = control	NR	No	No
53	No	Therapist	No		Component analysis			NR	No	No
54	Yes, no data available	Independent assessors	Yes	Yes	Evaluation of model	CT		One to four	No	No
55	No	Trained students	Yes	Yes	Comparison of therapy format			NR	No	Yes

continued

Study number*	Author response	Assessors	Blinded	Inter-rater reliability	Primary aim of trial	Models	Model weight	Adverse effects	Graphs	Subgroups
56	No	Therapist	No		Comparison of models/ fitness/self	BT/ST	Equal	NR	Yes	No
57	Yes, new data given	Experimenter	No		Component analysis			NR	No	No
58	Yes, new data given	NR			Evaluation of model	CBT		NR	No	No
59	No	NR			Comparison of therapy format	Individual/group		NR	No	No
60	No	Therapist	Co-raters only	Yes	Comparison of models	CT/BT	Equal	NR	No	No
61	Yes, no new data	Social skills measures/raters	Yes	No	Component analysis			NR	No	No
62	No	Evaluator	Yes	Yes	Component analysis			NR	No	No
63	No	Independent evaluator	Yes	Yes	Comparison of therapy format	Individual/group		NR	No	No

continued

Study number*	Variability measures	Comparable	Additional pharmacotherapy	Additional psychotherapy	Suicidal ideation	Number of sessions	Sample size	Number included in review	Total dropouts at post-treatment	Overall attrition
1	SDs	Yes	70% of total sample on psychotropic medication	Excluded if ≥ three sessions in 5 years		8 or 16	54	54	24 (18 + 6)	44%
2	No	Yes	NR	NR		7	20	20	5	20%
3	SDs	Does not appear to be	NR	NR		2	30	30	NR	
4	SDs	Yes	Both arms given pharmacotherapy placebo	Not on active treatment		12	125	64	17	26%
5	SDs	Yes	Withdrawn from medication pre-trial	Excluded if in therapy	Suicidal gestures	20	76	55	29	53%
6	SDs	Yes	35% of total sample on concurrent treatment	35% on concurrent treatment		12	80	63	NR	
7	SDs	NR	NR	NR		2	155	155	NR	
8	No	Yes	NR	NR	Severely suicidal	5	28	28	NR	
9	SDs	NR	NR	NR	Risk of suicide	2	57	57	NR	
10	No	NR	Imipramine/diazepam/ placebo arms	NR	Suicidal risk	3 + 15	218	NR	NR	
11	No	Yes	Imipramine arm excluded	NR		16	90	56	9	16%
12	No	Yes	Amitryptyline/combination arms excluded	Excluded if in therapy		16	96	48	29	60%
13	SDs	Yes	Imipramine arm excluded	NR	Suicidal potential	16	250	217	58	27%
14	SDs	Does not appear to be	NR	NR		2	49	49	1	2%
15	No	NR	NR	NR		8	35	35	5	14%
16	SDs	Yes	NR	Excluded if in therapy	Suicidal ideation	6	28	28	8	28%
17	No	Yes	Excluded if taking medication for depression	Excluded if in therapy	Suicidal risk	16–20	66	66	14	21%
18	No	NR	NR	NR		6	29	29	10	34%
19	One SD	NR	NR	NR		2	30	30	NR	

continued

Study number*	Variability measures	Comparable	Additional pharmacotherapy	Additional psychotherapy	Suicidal ideation	Number of sessions	Sample size	Number included in review	Total dropouts at post-treatment	Overall attrition
20	SDs	Yes	Excluded if taking medication for depression	NR		16–20	40	40	6	15%
21	Some SDs	Yes	NR	11 subjects were having concurrent counselling	Attempted suicide	8	28	28	2	7%
22	SDs	Yes	NR	Excluded if in therapy	Suicidal lethality	8	37	27	NR	
23	SDs	Yes	Excluded if on medication/in therapy for depression	Excluded if in therapy	Suicide potential	Up to 20	152	107	1	1%
24	SDs	NR	NR	NR		4	61	61	2	3%
25	SDs sent	Yes	Adherence to antidepressants was primary outcome	NR	Suicidal ideation	4–6	153	153	13	8%
26	SDs	Yes	Excluded if taking medication for depression	Excluded if in therapy	Suicidal crisis	12	49	49	10	20%
27	No	Does not appear to be	Three subjects on antidepressants	Excluded if in therapy	Suicidal tendencies	5	33	33	1	3%
28	Some SDs	Yes	Amitryptyline arm excluded	NR		10	196	143	14	9%
29	SDs	Yes	NR	NR	Suicidal behaviour	8	50	50	10	20%
30	SDs	Yes	57% taking antidepressants	34% in therapy		6	71	71	20	28%
31	SDs/SEs of the means	Male/female differences	Excluded if taking medication for depression	NR		20	37	25	3	12%
32	Pre-SDs	Yes	Excluded if on medication for depression	Excluded if in therapy		10	146	146?	NR	
33	SDs	Yes	Excluded if on medication for depression	Excluded if in therapy		8	32	32	6	19%
34	SDs	Yes	Excluded if on medication for depression	Excluded if in therapy		10	43	43	3	7%
35	SDs	BDI†	Excluded if on medication for depression	Excluded if in therapy	Suicidal ideation	Up to 8	89	89	4	4%

continued

Study number*	Variability measures	Comparable	Additional pharmacotherapy	Additional psychotherapy	Suicidal ideation	Number of sessions	Sample size	Number included in review	Total dropouts at post-treatment	Overall attrition
36	SDs	Yes	Sample group required to stop medication pre-trial	NR		12	24	24	1	2%
37	SDs	Yes	NR	Excluded if in therapy	Actively suicidal	7	24	24	NR	
38	SDs	Yes	Excluded if on medication for depression	Excluded if in therapy		10	58	58	7	12%
39	SDs	Yes	Excluded if on medication for depression	Excluded if in therapy		6	138	138	34	25%
40	SDs	NR	≥ 40 subjects on antidepressants	NR		12 or 15	67	67	19	28%
41	SDs	Education	None in individual CT and five in group CT given antidepressants	NR		Up to 20	44	37	6	16%
42	No	Yes	Ten subjects on medication in traditional arm, none in assertive arm	NR	Imminently suicidal	10	32	32	10	31%
43	SDs	Yes	NR	NR	Suicidal ideation	8	56	56	2	4%
44	SDs	Yes	45% of treatment as usual arm given antidepressants during trial	Excluded if in treatment for MDD		16 + 4	276	185	46 in IPT	25%
45	SDs/CIs	Does not appear to be	29 in treatment as usual group given antidepressants, none in CBT group, and number unknown in counselling group	NR	Suicidal risk	9–12	121	90	3	3%
46	SDs	Premorbid‡	23 of 24 subjects in each arm given antidepressants	NR		6	48	48	14	29%
47	SDs	NR	Just under half of the total sample given antidepressants	NR		12	36	36	3	8%
48	SDs	Yes	NR	NR		6	36	36	0	0%
49	SDs	Yes	25 of 117 subjects on medication for depression	Excluded if subjects had had three or more sessions in 5 years		8 or 16	150	150	27	18%

continued

Study number*	Variability measures	Comparable	Additional pharmacotherapy	Additional psychotherapy	Suicidal ideation	Number of sessions	Sample size	Number included in review	Total dropouts at post-treatment	Overall attrition
50	No	Yes	NR	NR	Suicidal behaviour	8	32	32	NR	
51	SDs	Yes	NR	NR		3	24	24	0	0%
52	SDs	NR	NR	NR		3	28	28	1	3%
53	SDs	Yes	Excluded if on medication for depression	Excluded if in therapy		6	28	28?	NR	
54	Medians	Does not appear to be	11 subjects in CT group, ten subjects in treatment as usual group	Excluded if in therapy		Up to 20	44	44	11	25%
55	SDs	Yes	NR	26% subjects in therapy		12	90	66	NR	
56	SDs	Yes	Excluded if on medication for depression	Excluded if in therapy		3	56	28	12–22%	
57	SDs	Yes	NR	NR	Risk of suicide	4	63	63	7?	24%
58	No	Yes	Minor tranquillisers only allowed	Excluded if in therapy		4	60	60	20	33%
59	SDs	Yes	Excluded if on medication for depression	Excluded if in therapy	Suicidal ideation	6	18	18	0	0%
60	SDs	NR	Two subjects on antidepressants, three on minor tranquillisers	NR	Suicidal ideation	8	25	25	4	16%
61	No	Yes	NR	Excluded if in therapy		12	66	66	22	33%
62	SDs	NR	Excluded if on antidepressants	NR		12	37	37	6	16%
63	SDs	HRSD§	Excluded if on antidepressants	NR		12	30	30	3	22%

* *See appendix 6 for the references of the study numbers*
† *There was a small but significant pretreatment difference between groups on BDI scores that was controlled for in the analyses using analysis of covariance*
‡ *The mean premorbid neuroticism scale score was significantly higher in the intervention group at baseline, but this was taken into account in analysing the outcome data*
§ *Significant pretreatment differences between groups on HRSD scores were controlled for in the analyses using analysis of covariance*

MAD, major affective disorder; NR, not reported; GP, general practitioner; RDC, Research Diagnostic Criteria; SADS, Schedule of Affective Disorders Scale

Appendix 3

Standard letter to authors

Direct Tel: 0171 919 3128
Fax: 0171 277 0283
Email address: v.hunot@iop.bpmf.ac.uk

Date: 17th December 1998

Dear

Reference:

A team of colleagues and I at the Institute of Psychiatry are currently conducting an **NHS R&D Health Technology Assessment funded Systematic Review of Brief Psychological Treatments for Depression**, which will provide a summary of the evidence for the clinical effectiveness and cost-effectiveness of these interventions. The completed review will be incorporated onto the Cochrane Library for worldwide dissemination, and all newly identified randomised controlled trials and controlled clinical trials will be included on the Cochrane Collaboration Depression Anxiety and Neurosis Clinical Trials Register (CCDANCTR). Within the UK, the results of the review will be made available to general practitioners via workshops and training packages, and an executive summary will be made available to NHS regional offices for dissemination within each region.

In order that the above study may be included in the review, we would be grateful for some further information from you, as indicated below.

In each comparison group:

- Number of subjects at randomisation ☐
- Dropout rate during intervention ☐
- Number of subjects at post-treatment assessment ☐
- Number of subjects at final follow-up assessment ☐
- Group mean depression scores and standard deviations at baseline ☐
- Group mean depression scores and standard deviations at post-treatment assessment ☐
- Group mean depression scores and standard deviations at final follow-up ☐
- Group mean change scores and standard deviations ☐
- Categorical/dichotomous outcomes (significant clinical improvement or recovery, defined as e.g. less than 10 on BDI) ☐

The trials we have identified so far are generally lacking in cost-effectiveness data. Therefore, we are also keen to locate studies (regardless of methodology) of time-limited psychotherapy for depression, which have a cost-effectiveness dimension.

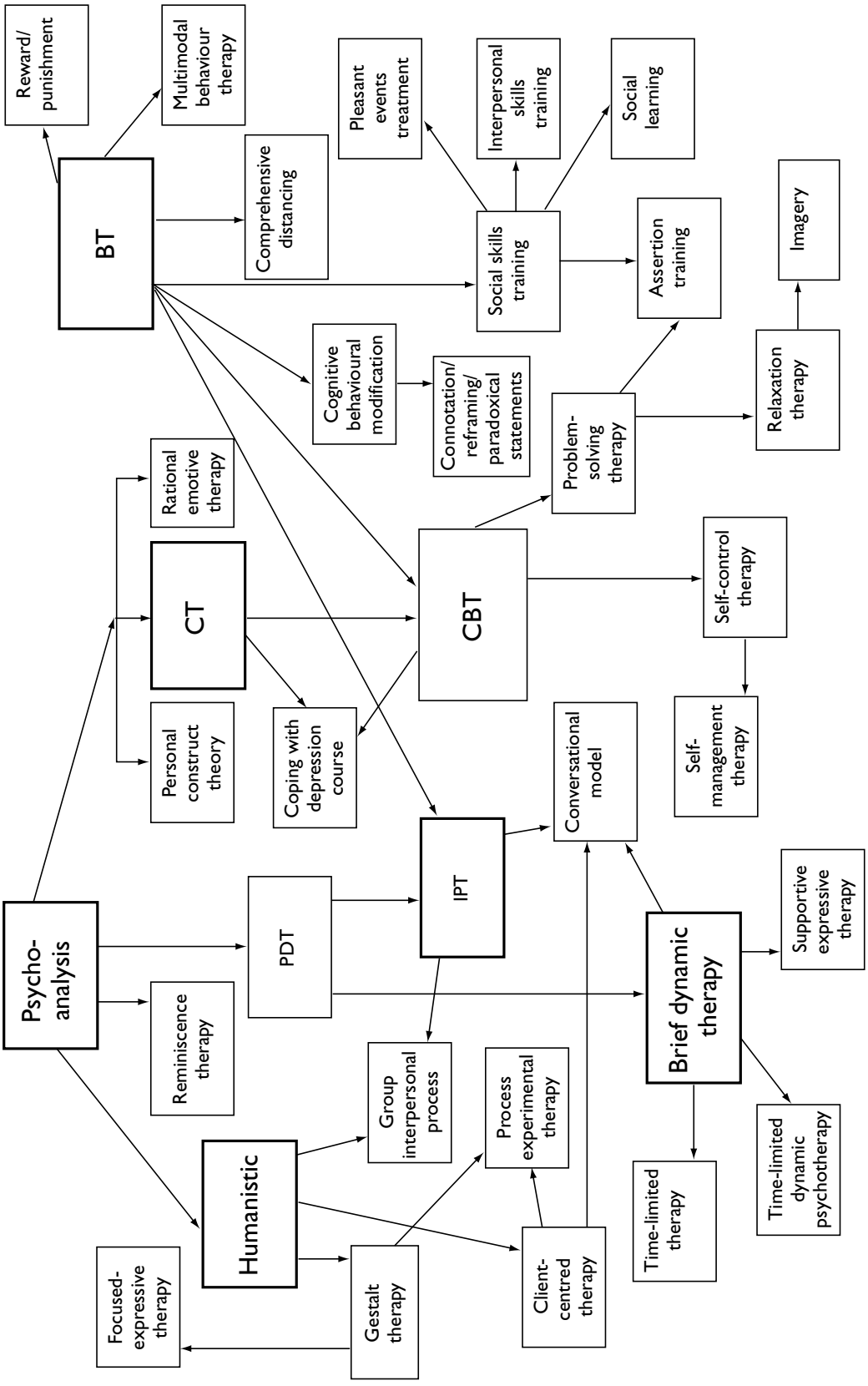
We would much appreciate any information that you are able to offer us, and your assistance will obviously be acknowledged in the completed review. **The deadline for returning data is 8th January 1999.** If you have any queries, please contact **Vivien Hunot on 0171 919 3128 or email v.hunot@iop.bpmf.ac.uk**.

Yours sincerely

Vivien Hunot, Research Worker
Rachel Churchill, Lecturer in Psychiatric Epidemiology
Martin Knapp, Professor of Economics of Mental Health
Roslyn Corney, Professor of Psychology at University of Greenwich

Appendix 4

Psychotherapy models


Reward/ punishment
Multimodal behaviour therapy
BT
Comprehensive distancing
Pleasant events treatment
Interpersonal skills training
Social learning
Social skills training
Assertion training
Imagery
Cognitive behavioural modification
Connotation/ reframing/ paradoxical statements
Problem-solving therapy
Relaxation therapy
Rational emotive therapy
CT
CBT
Self-control therapy
Personal construct theory
Coping with depression course
Conversational model
Self-management therapy
Psycho-analysis
PDT
IPT
Supportive expressive therapy
Reminiscence therapy
Brief dynamic therapy
Group interpersonal process
Process experimental therapy
Humanistic
Time-limited dynamic psychotherapy
Time-limited therapy
Focused-expressive therapy
Gestalt therapy
Client-centred therapy

Appendix 5

QRS instrument

Criteria	Categories
(1) Objectives and specification – main outcomes *a priori*	0 = objectives unclear 1 = objectives clear but main outcomes not specified *a priori* 2 = objectives clear with *a priori* specification of main outcomes
(2) Sample size (number per group)	0 = < 50 1 = 50–100 2 = > 100
(3) Planned duration of trial including follow-up	0 = < 3 months 1 = > 3 months–< 6 months 2 = > 6 months
(4) Power calculation	0 = not reported 1 = mentioned without details 2 = details of calculations provided
(5) Method of allocation	0 = not randomised and likely to be biased 1 = partially or quasi-randomised with some bias possible 2 = randomised allocation
(6) Concealment of allocation*	0 = not done or not reported 1 = partial concealment reported 2 = done adequately
(7) Clear description of treatment (including doses of drugs) and adjunctive treatment	0 = main treatments not clearly described 1 = inadequate details of main or adjunctive treatments 2 = full details of main and adjunctive treatments
(8) Blinding of subjects	0 = not done 1 = done but no test of blinding† 2 = done and integrity of blinding tested
(9) Source‡ of subjects described and representative sample recruitment	0 = source of subjects not described 1 = source of subjects but unrepresentative sample, e.g. volunteers 2 = source of subjects described plus representative sample taken
(10) Use of diagnostic criteria (or clear specification of inclusion criteria)	0 = none 1 = diagnostic criteria or clear inclusion criteria 2 = diagnostic criteria plus specification of severity
(11) Record of exclusion criteria and number of exclusions and refusals reported	0 = criteria and number not reported 1 = criteria or number of exclusions and refusals not reported 2 = criteria and number of exclusions and refusals reported
(12) Description of sample demographics	0 = little/no information (only age/sex) 1 = basic details (e.g. marital status/ethnicity) 2 = full description (e.g. socio-economic status/clinical history)
(13) Blinding of assessor	0 = not done 1 = done but no test of blinding 2 = done and integrity of blinding tested
(14) Assessment of compliance with experimental treatments (including attendance for therapy)	0 = not assessed 1 = assessed for some experimental treatments 2 = assessed for all experimental treatments

continued

Criteria	Categories
(15) Details on side-effects	0 = inadequate details 1 = recorded by group but details inadequate 2 = full side-effect profiles by group
(16) Record of number and reasons for withdrawal by group	0 = no information on withdrawals by group 1 = withdrawals by group reported without reason 2 = withdrawals and reason reported by group
(17) Outcome measures described clearly or use of validated instruments	0 = outcomes not described clearly 1 = some outcomes not clearly described 2 = outcomes described or valid and reliable instruments used
(18) Information on comparability and adjustment for differences in analysis	0 = no information on comparability 1 = some information on comparability with appropriate adjustment 2 = sufficient comparability information with appropriate adjustment
(19) Inclusion of withdrawals in analysis (ITT or endpoint)	0 = not included or not reported 1 = withdrawals included in analysis by estimation of outcome 2 = withdrawals followed up and included in analysis
(20) Presentation of results with inclusion of data for re-analysis of main outcomes (e.g. SDs)	0 = inadequate presentation 1 = adequate presentation 2 = comprehensive presentation
(21) Appropriate statistical analysis (including correction for multiple tests where applicable)	0 = inappropriate 1 = mainly appropriate 2 = appropriate and comprehensive
(22) Conclusions justified	0 = no 1 = partially 2 = yes
(23) Declaration of interests (e.g. source of funding)	0 = no 2 = yes

[*] *Details on how the allocation code was protected from those involved in patient recruitment. May be achieved by conducting allocation through a central independent body or protection of code by sealed opaque envelopes, for example*

[†] *Test of integrity of blinding is normally done by asking participants to guess their allocated group. Results can be compared to those that would be expected by chance*

[‡] *Source of subjects refers to the setting in which subjects were found, e.g. inpatients, outpatients, general practice, community, etc.*

Appendix 6

References for the studies included in the review

Study 1
Barkham M, Rees A, Shapiro DA, Stiles WB, Agnew RM, Halstead J, *et al.* Outcomes of time-limited psychotherapy in applied settings: replicating the Second Sheffield Psychotherapy Project. *J Consult Clin Psychol* 1996;**64**:1079–85.

Study 2
Barrera M. An evaluation of a brief group therapy for depression. *J Consult Clin Psychol* 1979;**47**:413–15.

Study 3
Beck JT, Strong SR. Stimulating therapeutic change with interpretations: a comparison of positive and negative connotation. *J Couns Psychol* 1982;**29**:551–9.

Study 4
Bellack AS, Hersen M, Himmelhoch JM. A comparison of social skills training, pharmacotherapy and psychotherapy for depression. *Behav Res Ther* 1983;**1**:101–7.

Hersen M, Bellack AS, Himmelhoch J, Thase ME. Effects of social skill training, amitriptyline and psychotherapy in unipolar depressed women. *Behav Ther* 1984;**15**:21–40.

Last CG, Thase ME, Hersen M, Bellack AS, Himmelhoch J. Patterns of attrition for psychosocial and pharmacologic treatments of depression. *J Clin Psychiatry* 1985;**46**:361–6.

Study 5
Beutler LE, Engle D, Mohr D, Daldrup RJ, Bergan J, Meredith K, *et al.* Predictors of differential response to cognitive, experiential and self-directed psychotherapeutic procedures. *J Consult Clin Psychol* 1991;**59**:333–40.

Study 6
Brown RA, Lewinsohn PM. A psychoeducational approach to the treatment of depression: comparison of group, individual and minimal contact procedures. *J Consult Clin Psychol* 1984;**52**:774–83.

Study 7
Catanese RA, Rosenthal TL, Kelley JE. Strange bedfellows: reward, punishment and impersonal distraction strategies in treating dysphoria. *Cogn Ther Res* 1979; **3**:299–305.

Study 8
Comas-Diaz L. Effects of cognitive and behavioral group treatment on the depressive symptomatology of Puerto rican women. *J Consult Clin Psychol* 1981;**49**:627–32.

Study 9
Conoley CW, Garber RA. Effects of reframing and self-control directives on loneliness, depression and controllability. *J Couns Psychol* 1985;**32**:139–42.

Study 10
Covi L, Lipman RS, Derogatis LR, Smith JE III, Pattison JH. Drugs and group psychotherapy in neurotic depression. *Am J Psychiatry* 1974;**131**:191–8.

Covi L, Lipman RS, Alarcon RD, Smith VK. Drug and psychotherapy interactions in depression. *Am J Psychiatry* 1976;**133**:502–8.

Study 11
Covi L, Lipman RS. Cognitive behavioral group psychotherapy combined with imipramine in major depression. *Psychopharmacol Bull* 1987;**23**:173–6.

Study 12
DiMascio A, Weissman MM, Prusoff BA, Neu C, Zwilling M, Klerman GL. Differential symptom reduction by drugs and psychotherapy in acute depression. *Arch Gen Psychiatry* 1979;**36**:1450–6.

Weissman MM, Prusoff B, DiMascio A, Neu C, Goklaney M, Klerman G. The efficacy of drugs and psychotherapy in the treatment of acute depressive episodes. *Am J Psychiatry* 1979;**136**:555–8.

Hercog-Baron RL, Prusoff B, Weissman MM, DiMascio A, Neu C, Klerman G. Pharmacotherapy and psychotherapy in acutely depressed patients: a study of attrition patterns in a clinical trial. *Compr Psychiatry* 1979;**20**:315–25.

Prusoff B, Weissman MM, Klerman G, Rounsaville BJ. Research diagnostic criteria subtypes of depression: their roles as predictors of differential response to psychotherapy and drug treatment. *Arch Gen Psychiatry* 1980;**37**:796–803.

Zuckerman DM, Prusoff B, Weissman MM, Padian N. Personality as a predictor of psychotherapy and pharmacotherapy outcome for depressed outpatients. *J Consult Clin Psychol* 1981;**48**:730–5.

Rounsaville BJ, Klerman G, Weissman MM. Do psychotherapy and pharmacotherapy for depression conflict? Empirical evidence from a clinical trial. *Arch Gen Psychiatry* 1981;**38**:24–9.

Weissman MM, Klerman G, Prusoff B, Shilomskas D, Padian N. Depressed outpatients: one year after treatment with drugs and/or interpersonal psychotherapy. *Arch Gen Psychiatry* 1981;**38**:51–5.

Study 13

Elkin I, Shea MT, Watkins JT, Imber SD, Sotsky SM, Collins JF, *et al.* National Institute of Mental Health Treatment of Depression Collaborative Research Program: general effectiveness of treatments. *Arch Gen Psychiatry* 1989;**46**:971–82.

Elkin I, Parloff MB, Hadley SW, Autry JH. NIMH Treatment of Depression Collaborative Research Program: background and research plan. *Arch Gen Psychiatry* 1985;**42**:305–16.

Elkin I, Shea MT, Watkins JT, Gibbons RD, Sotsky SM, Pilkonis PA, *et al.* Initial severity and differential treatment outcome in the National Institute of Mental Health Treatment of Depression Collaborative Research Program. *J Consult Clin Psychol* 1995;**63**:841–7.

Imber SD, Pilkonis PA, Sotsky SM, Watkins JT, Shea MT, Elkin I, *et al.* Mode-specific effects among three treatments for depression. *J Consult Clin Psychol* 1990;**58**:352–9.

Shea MT, Pilkonis PA, Beckham E, Collins JF, Elkin I, Sotsky SM, *et al.* Personality disorders and treatment outcome in the NIMH Treatment of Depression Collaborative Research Program. *Am J Psychiatry* 1990;**147**:711–18.

Sotsky SM, Glass DR, Shea MT, Pilkonis PA, Collins JF, Elkin I, *et al.* Patient predictors of response to psychotherapy and pharmacotherapy: findings in the NIMH Treatment of Depression Collaborative Research Program. *Am J Psychiatry* 1991;**148**:997–1008.

Shea MT, Elkin I, Imber SD, Sotsky SM, Watkins JT, Collins JF, *et al.* Course of depressive symptoms over follow-up: findings from the NIMH Treatment of Depression Collaborative Research Program. *Arch Gen Psychiatry* 1992;**49**:782–7.

Watkins JT, Leber WR, Imber SD, Collins JF, Elkin I, Pilkonsis PA, *et al.* Temporal course of change of depression. *J Consult Clin Psychol* 1993;**61**:858–64.

Blatt SJ, Quinlann DM, Pilkonis PA, Shea MT. Impact of perfectionism and need for approval on the brief treatment of depression: NIMH Treatment of Depression Collaborative Research Program revisited. *J Consult Clin Psychol* 1995;**63**:125–32.

Blatt SJ, Quinlan DM, Suroff DC, Pilkonis PA. Interpersonal factors in brief treatment of depression: further analyses of the NIMH Treatment of Depression Collaborative Research Program. *J Consult Clin Psychol* 1996;**64**:162–71.

Krupnick JL, Sotsky SM, Simmens S, Moyer J, Elkin I, Watkins J, *et al.* The role of the therapeutic alliance in psychotherapy and pharmacotherapy outcome: findings in the National Institute of Mental Health Treatment of Depression Collaborative Research Program. *J Consult Clin Psychol* 1996;**64**:532–9.

Zlotnick C, Shea MT, Pilkonis PA, Elkin I, Ryan C. Gender, type of treatment, dysfunctional attitudes, social support, life events and depressive symptoms over naturalistic follow-up. *Am J Psychiatry* 1996;**153**:1021–7.

Sotsky SM, Simmens S. Pharmacotherapy and psychotherapy response in atypical depression – findings from the NIMH Treatment of Depression Collaborative Research Program. Proceedings of 149th Annual Meeting of the American Psychological Association; 1996 May 4–9; New York, USA. New York: American Psychiatric Association; 1996, Vol 5.

Study 14

Feldman DA, Strong SR, Danser DB. A comparison of paradoxical and nonparadoxical interpretations and directives. *J Couns Psychol* 1982;**29**:572–9.

Study 15

Fleming BM, Thornton DW. Coping skills training as a component in the short-term treatment of depression. *J Consult Clin Psychol* 1980;**48**:652–4.

Study 16

Fuchs CZ, Rehm LP. A self-control behavior therapy program for depression. *J Consult Clin Psychol* 1977;**45**:206–15.

Study 17

Gallagher Thompson D, Steffen AM. Comparative effects of cognitive-behavioral and brief psychodynamic psychotherapies for depressed family caregivers. *J Consult Clin Psychol* 1994;**62**:543–9.

Study 18

Gardner P, Oei TP. Depression and self-esteem: an investigation that used behavioral and cognitive approaches for the treatment of clinically depressed clients. *J Clin Psychol* 1981;**37**:128–35.

Study 19

Gold SR, Jarvinen PJ, Teague RG. Imagery elaboration and clarity in modifying college students' depression. *J Clin Psychol* 1982;**38**:312–14.

Study 20

Greenberg LS, Watson J. Experiential therapy of depression: differential effects of client-centered relationship conditions and process experiential interventions. *Psychother Res* 1998;**8**:210–24.

Study 21
Hayman PM, Cope CS. Effects of assertion training on depression. *J Clin Psychol* 1980;**36**:534–43.

Study 22
Hogg JA, Deffenbacher JL. A comparison of cognitive and interpersonal-process group therapies in the treatment of depression among college students. *J Couns Psychol* 1988;**35**:304–10.

Study 23
Jacobson NS, Dobson KS, Truax PA, Addis ME, Koerner K, Gollan JK, *et al.* A component analysis of cognitive-behavioral treatment for depression. *J Consult Clin Psychol* 1996;**64**:295–304.

Study 24
Jarvinen PJ, Gold SR. Imagery as an aid in reducing depression. *J Clin Psychol* 1981;**37**:523–9.

Study 25
Katon W, Robinson P, Von Korff M, Lin E, Bush T, Ludman E, *et al.* A multi-faceted intervention to improve treatment of depression in primary care. *Arch Gen Psychiatry* 1996;**53**:924–32.

Simon GE, Katon W, Rutter C, VonKorff M, Kin E, Robinson P, *et al.* Impact of improved depression treatment in primary care on daily functioning and disability. *Psychol Med* 1998;**28**:693–701.

VonKorff M, Katon W, Bush T, Lin EH, Simon GE, Saunders K, *et al.* Treatment costs, cost offset, and cost-effectiveness of collaborative management of depression. *Psychosom Med* 1998;**60**:143–9.

Study 26
Kornblith S, Rehm L, O'Hara MW, Lamparski DM. The contribution of self-reinforcement training and behavioral assignments to the efficacy of self-control therapy for depression. *Cogn Ther Res* 1983;**7**:499–528.

Study 27
LaPointe KA, Rimm DC. Cognitive, assertive and insight-oriented group therapies in the treatment of reactive depression in women. *Psychother: Theor Res Prac* 1980;**17**:312–21.

Study 28
Mclean PD, Hakstian AR Clinical depression: comparative efficacy of outpatient treatments. *J Consult Clin Psychol* 1979;**47**:818–36.

McLean PD, Hakstian AR. Relative endurance of unipolar depression treatment effects: longitudinal follow-up. *J Consult Clin Psychol* 1990;**58**:482–8.

McLean PD, Taylor S. Severity of unipolar depression and choice of treatment. *Behav Res Ther* 1992;**30**:443–51.

Taylor S, McLean PD. Outcome profiles in the treatment of unipolar depression. *Behav Res Ther* 1993;**31**:325–30.

Study 29
McNamara K, Horan JJ. Experimental construct validity in the evaluation of cognitive and behavioral treatments for depression. *J Couns Psychol* 1986;**33**:23–30.

Study 30
Morris NE. A group self-instruction method for the treatment of depressed outpatients [dissertation]. Toronto, Canada: University of Toronto; 1975.

Study 31
Murphy GE, Carney RM, Knesevich MA, Wetsel RD, Whitworth P. Cognitive behavior therapy, relaxation training and tricyclic antidepressant medication in the treatment of depression. *Psychol Rep* 1995;**77**:403–20.

Study 32
Neimeyer RA, Weiss ME. Cognitive and symptomatic predictors of outcome of group therapies for depression. *J Cogn Psychother* 1990;**4**:23–32.

Study 33
Nezu AM. Efficacy of a social problem-solving therapy approach for unipolar depression. *J Consult Clin Psychol* 1986;**54**:196–202.

Study 34
Nezu AM, Perri MG. Social problem-solving therapy for unipolar depression: an initial dismantling investigation. *J Consult Clin Psychol* 1989;**57**:408–13.

Study 35
Pace TM, Dixon DN. Changes in depressive self-schemata and depressive symptoms following cognitive therapy. *J Couns Psychol* 1993;**23**:209–14.

Study 36
Padfield M. The comparative effects of two counseling approaches on the intensity of depression among rural women of low socioeconomic status. *J Couns Psychol* 1976;**23**:209–14.

Study 37
Rehm L, Fuchs CZ, Roth DM, Kornblity SJ, Romano JM. A comparison of self-control and assertion skills treatments of depression. *Behav Ther* 1979;**10**:429–42.

Study 38
Rehm LP, Kornblith SJ, O'Hara MW, Lamparski DM, Romano JM, Volkin JI. An evaluation of major components in a self-control therapy program for depression. *Behav Modif* 1981;**5**:459–89.

Study 39
Rehm LP, Kaslow NJ, Rabin AS. Cognitive and behavioral targets in a self-control therapy program for depression. *J Consult Clin Psychol* 1987;**55**:60–7.

Study 40
Ross M, Scott M. An evaluation of the effectiveness of individual and group cognitive therapy in the treatment of depressed patients in an inner city health center. *J R Coll Gen Pract* 1985;**35**:239–42.

Scott M, Stradling SG. Group cognitive therapy for depression produces clinically significant reliable change in community-based settings. *Behav Psychother* 1990;**18**:1–19.

Study 41
Rush AJ, Watkins JT. Group versus individual cognitive therapy: a pilot study. *Cogn Ther Res* 1981;**5**:95–103.

Study 42
Sanchez VC, Lewinsohn PM, Larson DW. Assertion training: effectiveness in the treatment of depression. *J Clin Psychol* 1980;**36**:526–9.

Study 43
Schmidt MM, Miller WR. Amount of therapist contact and outcome in a multi-dimensional treatment program. *Acta Psychiatr Scand* 1983;**67**:319–32.

Study 44
Schulberg HC, Block MR, Madonia MJ, Scott CP, Rodriguez E, Imber SD, *et al.* Treating major depression in primary care practice. Eight month clinical outcomes. *Arch Gen Psychiatry* 1996;**53**:913–19.

Lave JR, Frank RG, Schulberg HC, Kamlet MS. Cost-effectiveness of treatments for major depression in primary care practice. *Arch Gen Psychiatry* 1998; **55**:645–51.

Study 45
Scott AI, Freeman CP. Edinburgh primary care depression study: treatment outcome, patient satisfaction and cost after 16 weeks. *BMJ* 1992;**304**:883–7.

Patience DA, McGuire RJ, Scott AI, Freeman CP. The Edinburgh Primary Care Depression Study: personality disorder and outcome. *Br J Psychiatry* 1995;**167**:324–30.

Study 46
Scott C, Tacchi MJ, Jones R, Scott J. Acute and one-year outcome of a randomised controlled trial of brief cognitive therapy for major depressive disorder in primary care. *Br J Psychiatry* 1997;**171**:131–4.

Study 47
Scott M, Stradling SG. Group cognitive therapy for depression produces clinically significant reliable change in community-based settings. *Behav Psychother* 1990;**18**:1–19.

Study 48
Selmi PM, Klein MH, Greist JH, Sorrell SP, Erdman HP. Computer-administered cognitive-behavioral therapy for depression. *Am J Psychiatry* 1990;**147**:51–60.

Study 49
Shapiro DA, Barkham M, Rees A, Hardy GE, Reynolds S, Startup M. Effects of treatment duration and severity of depression on the effectiveness of cognitive-behavioral and psychodynamic-interpersonal therapy. *J Consult Clin Psychol* 1994;**62**:522–34.

Shapiro DA, Barkham M, Hardy GE, Morrison LA. The Second Sheffield Psychotherapy Project: rationale, design and preliminary outcome data. *Br J Med Psychol* 1990;**63**:97–108.

Startup M, Shapiro DA. Therapist treatment fidelity in prescriptive versus exploratory psychotherapy. *Br J Clin Psychol* 1993;**32**:443–56.

Field SD, Barkham M, Shapiro DA, Stiles WB. Assessment of assimilation in psychotherapy: a quantitative case study of problematic experiences with a significant other. *J Couns Psychol* 1994;**41**:397–406.

Shapiro DA, Barkham M, Rees A, Hardy GE, Reynolds S, Startup M. Looking as strongly as we should in the right places: reply to Norcross and Rossi (1994). *J Consult Clin Psychol* 1994;**62**:539–42.

Stiles WB, Reynolds S, Hardy GE, Rees A, Barkham M, Shapiro DA. Evaluation and description of psychotherapy sessions by clients using the Session Evaluation Questionnaire and the Session Impacts Scale. *J Couns Psychol* 1994;**41**:175–85.

Shapiro DA, Rees A, Barkham M, Hardy G, Reynolds S, Startup M. Effects of treatment duration and severity of depression on the maintenance of gains after cognitive-behavioral and psychodynamic-interpersonal psychotherapy. *J Consult Clin Psychol* 1995;**63**:378–87.

Hardy GE, Barkham M, Shapiro DA, Reynolds S, Rees A. Credibility and outcome of cognitive-behavioural and psychodynamic-interpersonal psychotherapy. *Br J Clin Psychol* 1995;**34**:555–69.

Hardy GE, Barkham M, Shapiro DA, Stiles WB, Rees A, Reynolds S. Impact of Cluster C personality disorders on outcomes of contrasting brief psychotherapies for depression. *J Consult Clin Psychol* 1995;**63**:997–1004.

Reynolds S, Stiles WB, Barkham M, Shapiro DA, Hardy GE, Rees A. Acceleration of changes in session impact during contrasting time-limited psychotherapies. *J Consult Clin Psychol* 1996;**64**:577–86.

Study 50
Shaw BF. Comparison of cognitive therapy and behavior therapy in the treatment of depression. *J Consult Clin Psychol* 1977;**45**:543–51.

Study 51/52
Shipley CR, Fazio AF. Pilot study of a treatment for psychological depression. *J Abnorm Psychol* 1973;**82**:372–6.

Study 53
Taylor FG, Marshall WL. Experimental analysis of a cognitive-behavioral therapy for depression. *Cogn Ther Res* 1977;**1**:59–72.

Study 54
Teasdale JD, Fennell MJ, Hibbert GA, Amies PL. Cognitive therapy for major depressive disorder in primary care. *Br J Psychiatry* 1984;**144**:400–6.

Fennell MJ, Teasdale JD. Effects of distraction on thinking and affect in depressed patients. *Br J Clin Psychol* 1984;**23**:65–6.

Study 55
Teri L, Lewinsohn PM. Individual and group treatment of unipolar depression: comparison of treatment outcome and identification of predictors of successful treatment outcome. *Behav Ther* 1986;**17**:215–28.

Study 56
Turner RW, Ward MF, Turner DJ. Behavioral treatment for depression: an evaluation of therapeutic components. *J Clin Psychol* 1979;**35**:166–75.

Study 57
Tyson GM, Range LM. Gestalt dialogues as a treatment for mild depression: time works just as well. *J Clin Psychol* 1987;**43**:227–31.

Study 58
Usaf SO, Kavanagh DJ. Mechanisms of improvement in treatment for depression: test of a self-efficacy and performance model. *J Cogn Psychother* 1990;**4**:51–70.

Study 59
Wierzbicki M, Barlett TS. The efficacy of group and individual cognitive therapy for mild depression. *Cogn Ther Res* 1987;**11**:337–42.

Study 60
Wilson PH, Goldin JC, Charbonneau Powis M. Comparative efficacy of behavioral and cognitive treatments of depression. *Cogn Ther Res* 1983;**7**:111–24.

Study 61
Zeiss AM, Lewinsohn PM, Munoz RF. Nonspecific improvement effects in depression using interpersonal skills training, pleasant activity schedules or cognitive training. *J Consult Clin Psychol* 1979;**47**:427–39.

Study 62
Zettle RD, Haflich JL, Reynolds RA. Responsivity of cognitive therapy as a function of treatment format and client personality dimensions. *J Clin Psychol* 1992;**48**:787–97.

Study 63
Zettle RD, Rains JC. Group cognitive and contextual therapies in treatment of depression. *J Clin Psychol* 1989;**45**:436–45.

Appendix 7

References for the studies excluded from the review and the reason for exclusion

Reference of study	Reason for exclusion
Barkham M, Shapiro DA, Hardy GE, Rees A. Psychotherapy in two-plus-one sessions: outcomes of a randomized controlled trial of cognitive-behavioral and psychodynamic-interpersonal therapy for subsyndromal depression. *J Consult Clin Psychol* 1999;**67**:201–11	Study included subjects below diagnostic threshold for depression
Dunn RJ. Cognitive modification with depression-prone psychiatric patients. *Cogn Ther Res* 1979;**3**:307–17	Waiting-list control group given antidepressant medication
Fremont J, Craighead LW. Aerobic exercise and cognitive therapy in the treatment of dysphoric moods. *Cogn Ther Res* 1987;**11**:241–52	Control group given aerobic exercise intervention
Graff RW, Whitehead GI, LeCompte M. Group treatment with divorced women using cognitive-behavioral and supportive-insight methods. *J Couns Psychol* 1986;**33**:276–81	Primary inclusion criterion defined as divorce – subjects were not required to be depressed for inclusion in study
Hammen CL, Glass DR. Depression, activity and evaluation of reinforcement. *J Abnorm Psychol* 1975;**84**:718–21	Therapeutic intervention self-administered by subjects
Heiby E. Social versus self-control deficits in four cases of depression. *Behav Ther* 1986;**17**:158–69	Crossover design utilised
Katon W, Von Korff M, Lin E, Walker E, Simon GE, Bush T, *et al*. Collaborative management to achieve treatment guidelines: impact on depression in primary care. *JAMA* 1995;**273**:1026–31	Description of psychological intervention was an educational package
Miranda J, Munoz R. Intervention for minor depression in primary care patients. *Psychosom Med* 1994;**56**:136–41	Only one-third of the sample group met the inclusion criteria for depression
Neimeyer RA, Neixas G. The role of homework and skill acquisition in the outcome of group cognitive therapy for depression. *Behav Ther* 1990;**21**:281–92	Dismantling study of one psychological treatment only
Rude SS. Relative benefits of assertion or cognitive self-control treatment for depression as a function of proficiency in each domain. *J Consult Clin Psychol* 1986;**54**:390–4	Crossover design utilised
Thase ME, Reynolds CF, Frank E, Simons AD, Garamoni GD, McGeary J, *et al*. Response to cognitive-behavioral therapy in chronic depression. *J Psychother Practice Res* 1994;**3**:204–14	Comparison between chronic depression versus dysthymia using same psychological intervention for both groups
Wilson GL. Psychotherapy with depressed incarcerated felons: a comparative evaluation of treatments. *Psychol Rep* 1990;**67**:1027–41	Sample group taken from a prison population and thus not representative
Zettle RD, Herring EL. Treatment utility of the sociotropy/autonomy distinction: implications for cognitive therapy. *J Clin Psychol* 1995;**51**:280–8	Randomisation procedure manipulated by matching and mismatching subjects to treatment according to personality dimensions

Appendix 8

Characteristics of included studies

Study number*	Methods	Inclusion criteria of participants	Interventions	Outcomes measures†	Allocation
1	*Method of randomisation*: stratified according to severity of depression levels *Allocation concealment*: not reported *Analysis*: completers only	Non-psychotic MDD (DSM-III), Present State Examination Index of ≥ 5, BDI score of ≥ 16, SCL-90-revised	CBT (eight sessions) versus PDT/IPT (eight sessions) CBT (16 sessions) versus PDT/IPT (16 sessions)	**BDI, SCL-90-revised**, Inventory of Interpersonal Problems and Self-esteem Scale	D
2	*Method of randomisation*: not described *Allocation concealment*: none *Analysis*: completers only	Clinical depression (as defined by the MMPI) and interview ratings used by Lewinsohn *et al.*[65]	BT versus a waiting-list control	**MMPI, BDI**, PES, Mood Related Items Scale (PES subscale) and LDACL[66]	D
3	*Method of randomisation*: not described *Allocation concealment*: none *Analysis*: not reported	Negative mood state and a BDI score of ≥ 7	Positive connotation (Palazzoli *et al.*[67]) versus a waiting-list control Negative connotation (Beck[68] and Ellis[69]) versus a waiting-list control Positive connotation versus negative connotation	**BDI**, Personal Mood Inventory, Barrett–Lennard Relationship Inventory (40-item revised version) and a free response form completed	D
4	*Method of randomisation*: not described *Allocation concealment*: none *Analysis*: endpoint only	Non-psychotic, unipolar depression (as defined by the RDC), DSM-III MDD and a score of ≥ 7 on the Raskin Eligibility Depression Scale	Social skills therapy + placebo versus psychotherapy + placebo	**BDI, HRSD**, Raskin Eligibility Depression Scale, LDACL,[66] RGSD, Hopkins SCL-90, Eysenck Personality Inventory, Wolpe–Lazarus Assertiveness Scale, Social Adjustment Scale, behavioural data and dropouts	D
5	*Method of randomisation*: randomised in blocks of five to ten *Allocation concealment*: none *Analysis*: endpoint only	Non-bipolar, non-psychotic MDD (DSM-III), Diagnostic Interview Schedule validation and an HRSD score of ≥ 16	Experiential therapy versus CT	**BDI, HRSD, Global Severity Index**, Brief Symptom Inventory, Barrett–Lennard Relationship Inventory, Working Alliance Inventory and selected MMPI scales	D
6	*Method of randomisation*: not described *Allocation concealment*: none *Analysis*: completers only?	Unipolar major, minor or intermittent depression (as defined by the RDC and the SADS)	Individual BT versus a waiting-list control Group BT versus a waiting-list control Individual BT versus group BT Individual BT + group BT versus a waiting-list control	**BDI, Center for Epidemiologic Studies Depression Scale, Domains of Satisfaction Scale** and SADS – Change	D
7	*Method of randomisation*: subjects matched on scores and sex per condition *Allocation concealment*: none *Analysis*: completers only	Dysphoric symptomatology, a BDI score of ≥ 9 and a ZSRDS score of ≥ 40	Overt reward versus a waiting-list control Overt reward + covert reward + overt punishment + covert punishment + social influence versus a waiting-list control	**BDI** and **ZSRDS**	D
8	*Method of randomisation*: matched for age and severity of depression before random assignment to groups *Allocation concealment*: none *Analysis*: completers only	Non-psychotic depression as identified by local community agencies, the BDI and the HRSD were of unspecified level	CT versus a waiting-list control BT versus a waiting-list control CT versus BT CT + BT versus a waiting-list control	**BDI, HRSD** and Depression Behaviour Rating Scale completed by significant other	D
9	*Method of randomisation*: not described *Allocation concealment*: none *Analysis*: completers only	Symptoms of moderate depression and a BDI score of 8–26	Self-control therapy versus a waiting-list control Self-control therapy + reframing versus a waiting-list control	**BDI, Revised Loneliness Scale**, Causal Dimension Scale and Counsellor Rating Scale[70]	D

continued

Study number*	Methods	Inclusion criteria of participants	Interventions	Outcomes measures†	Allocation
10	*Method of randomisation*: not described *Allocation concealment*: none *Analysis*: endpoint and completers only	Mild/moderate neurotic depression and ≥ 7 on the RGSD	PDT versus ST	**Psychiatric Outpatients Mood Scale, Symptom Distress Checklist**, Global Improvement Scale and Brief Psychiatric Rating Scale	D
11	*Method of randomisation*: not described *Allocation concealment*: none *Analysis*: endpoint only	Primary major non-bipolar, non-psychotic depression (as defined by the RDC), a BDI score of ≥ 20 and ≥ 14 on the HRSD	CBT versus traditional psychotherapy	**BDI, HRSD**, Hopkins SCL and Global Improvement Scale	D
12	*Method of randomisation*: not described *Allocation concealment*: none *Analysis*: endpoint only	Non-bipolar, non-psychotic, acute primary MDD (as defined by the RDC) and ≥ 7 on the RGSD	IPT versus non-scheduled therapy	**RGSD, HRSD**, Hopkins SCL-90, Global Illness Rating, Social Adjustment Scale and dropouts	D
13	*Method of randomisation*: assignment was based on a separate computer-generated random order for each research site *Allocation concealment*: none *Analysis*: endpoint and completers only	Non-bipolar non-psychotic MDD (as defined by the RDC, the SADS and the SADS – Change) and an HRSD score of ≥ 14	CBT versus clinical management + placebo IPT versus clinical management + placebo CBT versus IPT CBT versus IPT + clinical management + placebo	**HRSD, Global Assessment Scale, BDI, Hopkins SCL-90 Total**, Dysfunctional Attitude Scale, Social Adjustment Scale and Longitudinal Interval Follow-up Evaluation II	D
14	*Method of randomisation*: stratified random procedure to ensure equivalent pre-test BDI means in each group *Allocation concealment*: none *Analysis*: completers only (but only one dropout)	Negative mood state, non-serious depression and a BDI score of 8–21	Paradoxical interpretation/paradoxical directive + non-paradoxical interpretation/non-paradoxical directive + non-paradoxical interpretation/paradoxical directive + paradoxical interpretation/non-paradoxical directive versus a no treatment control	**BDI, Mood Perception Inventory** and Barrett–Lennard's Relationship Inventory (revised)	D
15	*Method of randomisation*: not described *Allocation concealment*: none *Analysis*: completers only	Clinical judgement of non-psychotic clinical depression, ≥ 14 on the MMPI-D (unidimensional depression scale for MMPI) and a BDI score of ≥ 7	CT versus non-directive therapy BT versus non-directive therapy CT versus BT CT + BT versus non-directive therapy	**BDI, MMPI-D**, Dysfunctional Attitude Scale, Irrational Beliefs Scale, PES, Overall Depression Rating and Negative Self-reference Rate (not validated)	D
16	*Method of randomisation*: random assignment, except where necessary to balance experimental conditions for mean age and severity of depression *Allocation concealment*: none *Analysis*: completers only	Non-psychotic depression (as defined by the MMPI)	CBT versus a waiting-list control Non-specific therapy versus a waiting-list control CBT versus non-specific therapy CBT + non-specific therapy versus a waiting-list control	**MMPI, BDI**, PES, group interaction measure (not validated), self-evaluation rating of ideal self (not validated) and Concepts Test (not validated)	D
17	*Method of randomisation*: balance of short- and long-term care-giving histories and use of Efron's procedure ('biased coin' process) for biasing probable assignment in sequential experiment *Allocation concealment*: none *Analysis*: completers only	Non-bipolar, non-psychotic major, minor or intermittent depressive disorder (as defined by the RDC and the SADS) and a BDI score of ≥ 10	CBT versus PDT	**BDI, HRSD, Geriatric Depression Scale** and SADS – Change	D

continued

Study number*	Methods	Inclusion criteria of participants	Interventions	Outcomes measures†	Allocation
18	*Method of randomisation*: subjects stratified according to sex and depression levels *Allocation concealment*: none *Analysis*: completers only?	Depression (BDI of unspecified level) maintained over 3 weeks of baseline testings	CT versus BT	**BDI, ZSRDS** and Tennessee Self-concept	D
19	*Method of randomisation*: not described *Allocation concealment*: none *Analysis*: not reported – completers only?	Self-rated depression (as defined by a BDI score of ≥ 11)	Positive imagery therapy + positive imagery/ self-elaboration therapy versus a no treatment control	**BDI**, Visual Imagery Scale (from Imaginal Processes Inventory), Verbaliser–Visualiser Questionnaire[71] and Image Clarity Assessment (not validated)	D
20	*Method of randomisation*: clients matched on SCL-90-revised depression subscale scores *Allocation concealment*: none *Analysis*: completers only	Non-bipolar, non-psychotic MDD (Structured Clinical Interview for DSM-III-R72), SCL-90-revised depression subscale and a Global Assessment Scale score of ≥ 50	Process-experiential therapy versus client-centred therapy To be included in a subgroup analysis or for excluded studies list?	**BDI, SCL-90-revised**, Rosenberg Self-esteem Inventory, Inventory of Interpersonal Problems, Target Complaints Discomfort Box Scale,[73] Longitudinal Interval Follow-up Evaluation II, Working Alliance Inventory, Barrett–Lennard Perceived Empathy Scale from Barrett–Lennard Relationship Inventory and Truax Accurate Empathy Scale	D
21	*Method of randomisation*: subjects ranked by BDI scores, then 16 assigned to treatment groups by a randomised stratification procedure using a random numbers table. The remaining 12 subjects were assigned to the waiting-list control *Allocation concealment*: none *Analysis*: completers only	Mild/moderate range of depression (as defined by a BDI score of 13–30)	Assertion training versus a waiting-list control	**BDI, PES** and College Self-expression Scale	D
22	*Method of randomisation*: not described *Allocation concealment*: none *Analysis*: completers only? (numbers not reported)	Non-psychotic unipolar depression identified through two clinical interviews and a BDI score of ≥ 14	CT versus IPT (supportive)	**BDI, MMPI**, Automatic Thoughts Questionnaire, Self-esteem Inventory (adult form) and Therapist Assessment of Client Behaviours (not validated)	D
23	*Method of randomisation*: subjects matched on previous episodes of depression, presence/absence of dysthymia, severity of depression, gender and marital status *Allocation concealment*: none *Analysis*: ITT and endpoint analysis	Non-bipolar, non-psychotic MDD (Structured Clinical Interview for DSM-III-R72), a BDI score of ≥ 20 and an HRSD score of ≥ 14	CT versus BT	**HRSD, BDI** and **Longitudinal Interval Follow-up Evaluation**	D
24	*Method of randomisation*: not described *Allocation concealment*: none *Analysis*: completers only?	Self-rated depression as defined by a BDI score of ≥ 12	Neutral imagery + positive imagery versus a control	**BDI, ZSRDS** and Daily Mood Rating (referenced but not validated)	D

continued

Study number*	Methods	Inclusion criteria of participants	Interventions	Outcomes measures†	Allocation
25	*Method of randomisation*: patients stratified into moderate and severe depression groups, and were block randomised within each level (assignment equalised after every eighth assignment) using a computer-generated sequence *Allocation concealment*: none *Analysis*: ITT	Non-psychotic major depression (as defined by the SCL-20 adapted from the SCL-90)	CBT versus treatment as usual	**BDI, SCL-20, Inventory for Depressive Symptomatology**, NEO Personality Inventory,[74] Chronic Disease Score, adherence to medication and satisfaction with management of depression	D
26	*Method of randomisation*: groups were started as soon as a sufficient number of subjects could meet at a particular time *Allocation concealment*: none *Analysis*: completers only	Primary non-psychotic major affective disorder (as defined by the SADS and the RDC) and a BDI score of ≥ 20	Self-control therapy versus PDT Self-control therapy + self-monitoring/ self-evaluation + principles only self-control therapy versus PDT	**BDI, MMPI, HRSD**, RGSD, Global Assessment Scale, Self-control Questionnaire,[75] PES/Unpleasant Events Schedule, interpersonal functioning and overt-motor behaviour	D
27	*Method of randomisation*: groups formed in two consecutive 4-week blocks, and treatment conditions randomly assigned to groups *Allocation concealment*: none *Analysis*: completers only	Non-psychotic recurrent situationally related depression (as defined by a BDI score of 15–31)	CT versus insight-oriented therapy Assertion training versus insight-oriented therapy CT versus assertion training CT + assertion training versus insight-oriented therapy	**BDI**, Personality Data Form (Part 1),[76] Rathus Assertiveness Schedule, Self-rated Improvement Questionnaire (not validated) and Reference Person Questionnaire (sent to person who knew subject well, but not validated)	D
28	*Method of randomisation*: random assignment to groups and therapists *Allocation concealment*: none *Analysis*: dropouts were replaced (removal of replacements = completers only)	Primary diagnosis of non-bipolar non-psychotic depression (as defined by Feighner *et al.*[77]), an MMPI score of ≥ 25 for males and ≥ 29.5 for females, a BDI score of >= 23 and a score of ≥ 14 on the LDACL	BT versus PDT BT versus relaxation therapy PDT versus relaxation therapy BT versus PDT + relaxation therapy	**LDACL,**[66] **BDI** and **Eysenck Personality Questionnaire**	D
29	*Method of randomisation*: assignment in blocks of four to treatment groups *Allocation concealment*: none *Analysis*: completers only	Non-psychotic primary depression (as defined by a diagnostic clinical interview), a BDI score of ≥ 18 at intake and ≥ 16 at pre-test and an HRSD score of ≥ 20	CBT versus client-centred therapy CBT + CT + BT versus client-centred therapy CT versus BT CT versus client-centred therapy BT versus client-centred therapy CBT versus CT CBT versus BT	**BDI, HRSD, Automatic Thoughts Questionnaire, PES**, Cognitive Scale (not validated), Recalled Cognitions Exercise (not validated), Self-evaluated Social Skills (not validated), Behavioural Scale (not validated), Observer-evaluated Social Skills (not validated), client expectancy/satisfaction and adherence (counsellor rating)	D
30	*Method of randomisation*: not described *Allocation concealment*: none *Analysis*: completers only	Primary diagnosis of neurotic or reactive depression, no present symptomatology of major psycho-pathology and an unspecified BDI score	BT versus client-centred therapy BT versus a waiting-list control Client-centred therapy versus a waiting-list control	**BDI, ZSRDS**, Anxiety Scale (not validated), Self-esteem Reaction Scale, Behavioural Reaction Scale (not validated), Depression Consequences Scale, Satisfaction Scale (not validated), Treatment Effectiveness Scale (not validated), Modified Personal Orientation Inventory[78]	D

continued

Study number*	Methods	Inclusion criteria of participants	Interventions	Outcomes measures†	Allocation
31	*Method of randomisation*: conducted using a table of generated random numbers *Allocation concealment*: none *Analysis*: endpoint only (and data omitted from three subjects who failed to complete second HRSD)	Primary unipolar affective disorder (as defined by Feighner *et al.*[77]) and a score of ≥ 14 on the BDI and the HRSD	CBT versus relaxation training	**BDI, HRSD**, Self-control Scale,[79] MMPI, a battery of 15 clinical and self-report rating scales (only the scales pertinent to this review reported) and subjects' expectations/satisfaction of treatment	D
32	*Method of randomisation*: not described *Allocation concealment*: none *Analysis*: not described	Primary diagnosis of major unipolar depression (as defined by the RDC and Spitzer *et al.*[38]) and a BDI score of ≥ 16	CT + homework + CT + no homework versus IPT	**BDI, HRSD**, Attributional Style Questionnaire and Crandall Cognitions Inventory	D
33	*Method of randomisation*: not described *Allocation concealment*: none *Analysis*: completers only	Current episode of unipolar non-psychotic depression (as defined by the RDC and the SADS), a BDI score of ≥ 16 and an MMPI score of ≥ 71	Problem-solving (CBT) versus a waiting-list control Problem-focused (supportive) therapy versus a waiting-list control Problem-solving therapy versus problem-focused therapy Problem-solving therapy + problem-focused therapy versus a waiting-list control	**BDI, MMPI**, Problem-solving Inventory and Internal–External Locus of Control Scale	D
34	*Method of randomisation*: not described *Allocation concealment*: none *Analysis*: completers only	Non-psychotic unipolar MDD (as defined by the RDC), a BDI score of ≥ 20 and an HRSD score of ≥ 18	Problem-solving therapy versus a waiting-list control Problem-solving therapy + abbreviated problem-solving therapy versus a waiting-list control	**HRSD, BDI**, Problem-solving Inventory, Problem Checklist80 and evaluation of treatment rationale and therapist competence	D
35	*Method of randomisation*: random assignment to groups and counsellors *Allocation concealment*: none *Analysis*: completers only	Mild/moderate non-psychotic depression based on a BDI score of 10–29	CT versus a no treatment control	**BDI** and self-referent judgement procedure (after which recall accuracy and reaction time were measured)	D
36	*Method of randomisation*: coin flipping for every other subject in order of appearance *Allocation concealment*: none *Analysis*: all subjects completed assessment measures	Primary diagnosis of moderate depression (as defined by the Grinker Interview Checklist) and an unreported ZSRDS score	BT versus ST	**ZSRDS, Grinker Interview Checklist**, PES and LDACL[66]	D
37	*Method of randomisation*: not described *Allocation concealment*: none *Analysis*: completers only	Primary diagnosis of non-psychotic depression using clinical judgement based on MMPI profile/interview data	Self-control therapy versus assertion skills therapy	**MMPI, BDI**, PES, Common Associates Test, Concepts Test (not validated), Walpe–Lazarus Assertion Scale, overt-motor behaviour observation, self-evaluation questionnaire and overt assertion skill behaviour	D

continued

Study number*	Methods	Inclusion criteria of participants	Interventions	Outcomes measures†	Allocation
38	*Method of randomisation*: random assignment to conditions within constraints of time availability (conducted in two replications, i.e. cohort 2 received treatment when the treatment of cohort 1 was complete) *Allocation concealment*: None *Analysis*: not reported	Primary diagnosis of non-bipolar, non-psychotic major affective disorder (as defined by the RDC and the SADS), an MMPI score of ≥ 71 and an HRSD score (that was not reported)	Self-control CBT versus a waiting-list control Self-control therapy + self-monitoring techniques + self-monitoring/self-evaluation techniques + self-monitoring/self-reinforcement techniques versus a waiting-list control Self-control therapy versus other dismantling techniques as listed above	**MMPI, BDI, HRSD**, RGSD, Global Assessment Scale, PES, Unpleasant Events Schedule, Self-control Questionnaire[75] and direct observational behavioural measures	D
39	*Method of randomisation*: not clearly described as RCT in main paper, reference to random assignation made in secondary paper *Allocation concealment*: none *Analysis*: completers only	Non-psychotic, non-bipolar major affective disorder (as defined by the SADS and the RDC), a BDI score of ≥ 20 and an MMPI score of ≥ 70	CT versus BT CBT versus CT CBT versus BT	**BDI, MMPI, HRSD**, RGSD, Hopkins SCL, Brief Psychiatric Rating Scale, Social Adjustment Scale, Self-control Questionnaire,[75] Self-control Scale,[79] Dysfunctional Attitude Scale, PES, Unpleasant Events Schedule and Cognitive Events Schedule	D
40	*Method of randomisation*: predetermined sequential assignment *Allocation concealment*: none *Analysis*: ITT	Primary or probable MDD (as defined by the Present State Examination Index and the RDC) and a BDI score of ≥ 13	Individual CT versus a waiting-list control Group CT versus a waiting-list control Individual CT versus group CT Individual CT + group CT versus a waiting-list control	**Montgomery–Asberg Depression Scale, BDI** and Social Stress and Support Interview	D
41	*Method of randomisation*: although not strictly randomised, all assignments were determined prior to and independent of the evaluation of a specific subject *Allocation concealment*: none *Analysis*: completers only	Non-bipolar, non-psychotic definite depressive syndrome (as defined by Feighner *et al.*[77]), a BDI score of ≥ 20 and an HRSD score of ≥ 14	Individual CT versus group CT	**BDI**, **HRSD** and **MMPI**	D
42	*Method of randomisation*: not described *Allocation concealment*: none *Analysis*: completers only	Non-psychotic clinical depression (as defined by the Grinker Interview Checklist) and an MMPI score of ≥ 71	Assertion training versus PDT	**MMPI**, **BDI**, modifications of Index of General Affect, modifications of Domains of Satisfaction Scale and Assertion Inventory	D
43	*Method of randomisation*: not described *Allocation concealment*: partial *Analysis*: no dropouts (post-treatment assessment)	Non-bipolar, non-psychotic primary diagnosis of depression (as defined by a semi-structured diagnostic interview) and a BDI score of ≥ 10	Individual CBT versus a waiting-list control Small group CBT versus a waiting-list control Individual CBT versus small group CBT + large group CBT Individual CBT + small group CBT + large group CBT versus a waiting-list control	**BDI, MMPI, ZSRDS**, Rotter Internal–External Locus of Control Scale, Social Readjustment Rating Scale, Profile of Mood States, LDACL,[66] Beck Hopelessness Scale, Trail-making Test (psychomotor speed) and PES	D
44	*Method of randomisation*: not described *Allocation concealment*: none *Analysis*: ITT and completers only	MDD (Diagnostic Interview Schedule/ DSM-III-R) and an HRSD score of ≥ 13	IPT versus treatment as usual	**HRSD** and **BDI**	D
45	*Method of randomisation*: system of sealed envelopes. Once groups were full in any one treatment, the coordinator opened one or more further envelopes until one of the other treatment arms came up *Allocation concealment*: partial *Analysis*: ITT	Non-psychotic MDD (DSM-III-R)	CBT versus social work counselling CBT versus treatment as usual (general practitioner management) Social work counselling versus general practitioner management CBT + social work counselling versus general practitioner management	**HRSD**, Personality Assessment Schedule, Social Functioning Scale and evaluation of treatment questionnaire	D

continued

Study number*	Methods	Inclusion criteria of participants	Interventions	Outcomes measures†	Allocation
46	*Method of randomisation*: stratification according to gender and severity and chronicity of depression *Allocation concealment*: none *Analysis*: completers only	Non-bipolar, non-psychotic primary MDD (DSM-III-R) and a BDI score of ≥ 20	CT versus usual general practitioner care	**HRSD** and **BDI**	D
47	*Method of randomisation*: predetermined sequential allocation *Allocation concealment*: none *Analysis*: ITT and completers only	Primary diagnosis of MDD (as defined by the RDC) and a BDI score of ≥ 20	Individual CT versus group CT	**BDI** and Irritability, Depression Anxiety Questionnaire (depression subscale)	D
48	*Method of randomisation*: not described *Allocation concealment*: none *Analysis*: no dropouts at post-treatment assessment	Non-psychotic major, minor or intermittent depressive disorder (as defined by the RDC and the SADS), a BDI score of ≥ 16 and ≥ the 65th percentile on the SCL-90-revised (13 items)	CBT versus a waiting-list control CBT versus computer CBT (supportive) Computer CBT versus a waiting-list control CBT + computer CBT versus a waiting-list control	**BDI**, **HRSD**, SCL-90-revised and Automatic Thoughts Questionnaire	D
49	*Method of randomisation*: conducted by research interviewer working to a protocol specifying randomisation within severity groups and balancing client gender as far as possible *Allocation concealment*: partial *Analysis*: completers only	Non-psychotic MDD (DSM-III-R) and a BDI score of ≥ 17	CBT versus PDT/IPT	**BDI**, **SCL-90-revised** (depression and self-esteem subscales), Global Severity Index and Social Adjustment Scale (social subscale)	D
50	*Method of randomisation*: groups matched according to age and severity of depression *Allocation concealment*: none *Analysis*: not described	Primary diagnosis of non-psychotic clinical depression by interviewer, an HRSD score of ≥ 20, a Visual Analogue Scale score of ≥ 40 and a BDI score of ≥ 18	CT + BT versus a waiting-list control CT versus BT CT + BT versus non-directive therapy CT + BT + non-directive therapy versus a waiting-list control	**BDI, HRSD, Visual Analogue Scale** and subject ratings of therapy/therapists	D
51	*Method of randomisation*: subjects were divided into two groups matched for age and severity of depression *Allocation concealment*: none *Analysis*: all subjects completed treatment	Primary self-rated depressive symptomatology (as defined by a score of ≥ 45 on the ZSRDS and ≥ 70 on the MMPI)	BT versus a waiting-list control	**MMPI** and **ZSRDS**	D
52	*Method of randomisation*: groups matched for age and depression levels *Allocation concealment*: none *Analysis*: not reported	Primary self-rated depressive symptomatology (as defined by a score of ≥ 45 on the ZSRDS and ≥ 70 on the MMPI)	BT + expectancy manipulation versus interest support therapy + expectancy manipulation BT versus interest support therapy	**MMPI** and **ZSRDS**	D
53	*Method of randomisation*: not described *Allocation concealment*: none *Analysis*: not described	Primary self-measured mild/moderate depression	CBT versus a waiting-list control CBT versus CT CBT versus BT CBT + CT + BT versus a waiting-list control	**BDI, MMPI-D**, Visual Analogue Scale, a version of Kelly's Repertory Grid, Eysenck Personality Inventory and Multiple Affect Adjective Checklist	D
54	*Method of randomisation*: not described *Allocation concealment*: none *Analysis*: completers only	Primary unipolar non-psychotic MDD (as defined by the RDC) and a BDI score of ≥ 20	CBT versus treatment as usual	**BDI, Montgomery–Asberg Depression Scale** and **HRSD**	D

continued

Study number*	Methods	Inclusion criteria of participants	Interventions	Outcomes measures†	Allocation
55	*Method of randomisation*: about two subjects assigned to group therapy to every one assigned to individual therapy *Allocation concealment*: none *Analysis*: completers only	Primary non-bipolar MDD or minor depressive disorder (as defined by the RDC)	Individual BT versus group BT	**SADS – Change, BDI**, Social Adjustment Scale, Perceived Social Support Inventory, Interpersonal Dependency Scale, Bem Sex-Role Inventory, Social Readjustment Rating Scale, life satisfaction and expectancy of treatment	D
56	*Method of randomisation*: simple rotation *Allocation concealment*: none *Analysis*: dropouts replaced	Primary depression (as defined by a score of ≥ 70 on the LDACL)	BT versus client-centred therapy	**LDACL,[80] Rotter Internal–External Locus of Control Scale** and PES	D
57	*Method of randomisation*: not described *Allocation concealment*: none *Analysis*: completers only	Non-psychotic mild depression (as defined by an MMPI score of ≥ 60)	Gestalt therapy versus a control	**MMPI, LDACL[80]** and client satisfaction (non-validated questionnaire)	D
58	*Method of randomisation*: subjects matched for age and BDI score *Allocation concealment*: none *Analysis*: completers only	MDD (as defined by the RDC), a BDI score of ≥ 18 and a ZSRDS score of ≥ 40	CBT versus a waiting-list control	**BDI, ZSRDS**, Self-efficacy Scale (non-validated), Automatic Thoughts Questionnaire, self-monitoring scales and dropout rate	D
59	*Method of randomisation*: not described *Allocation concealment*: none *Analysis*: all subjects included in *analysis*	Non-psychotic atypical affective disorder (DSM-III) and a BDI score of 8–35	Individual CT versus group CT	**BDI, MMPI-D (Dempsey)**, State–Trait Anxiety Inventory, Brief Symptom Inventory and Tennessee Self-concept	D
60	*Method of randomisation*: not described *Allocation concealment*: none *Analysis*: dropouts were replaced	Primary non-bipolar non-psychotic depressive symptomatology (as defined by a BDI score of ≥ 17)	CT versus BT CT + BT versus a waiting-list control	**BDI, HRSD**, Irrational Beliefs Test, PES (shortened version), Daily Mood Ratings (not validated) and Cognition Schedule (not validated)	D
61	*Method of randomisation*: not described *Allocation concealment*: none *Analysis*: completers only	Primary non-bipolar clinical depression (as defined by the Grinker Interview Checklist) and an MMPI score of ≥ 70	CT versus BT CT + BT versus a waiting-list control	**MMPI**, Interpersonal Events Schedule, PES, Cognitive Events Schedule, Personal Beliefs Inventory, Subjective Probability Questionnaire and interpersonal behaviour rated by coder and peers	D
62	*Method of randomisation*: not described *Allocation concealment*: none *Analysis*: completers only	Primary moderate–severe clinical depression (as defined by an MMPI score of ≥ 70), a BDI score of ≥ 20 and an HRSD score of ≥ 14	CT versus BT	**BDI, HRSD**, Automatic Thoughts Questionnaire, Dysfunctional Attitude Scale and PES	D
63	*Method of randomisation*: five male subjects treated individually in order to maintain homogeneous groups in terms of gender *Allocation concealment*: none *Analysis*: completers only	Primary moderate/severe clinical depression (as defined by the SADS), an MMPI score of ≥ 70, a BDI score of ≥ 20 and an HRSD score of ≥ 14	Individual CT versus group CT	**BDI, HRSD, MMPI**, Automatic Thoughts Questionnaire, Dysfunctional Attitude Scale, Sociotrophy Autonomy Scale (non-validated) and PES	D

* *See appendix 6 for the references for the studies*
† *The main outcome measures are shown in bold text*
MMPI-D, unidimensional depression scale for MMPI

Appendix 9

List of outcome instruments used

Outcome measures	Number of trials
Assertion Inventory	1
Attributional Style Questionnaire	1
Automatic Thoughts Questionnaire	5
Barrett–Lennard Relationship Inventory	4
Barrett–Lennard Relationship Inventory (40-item revised version)	1
BDI	54
Beck Hopelessness Scale	1
Bem Sex-role Inventory	1
Brief Psychiatric Rating Scale	2
Brief Symptom Inventory	2
Causal Dimension Scale	1
Center for Epidemiologic Studies Depression Scale	1
Chronic Disease Score	1
Cognitive Events Schedule	2
College Self-expression Scale	1
Common Associates Test	1
Counsellor Rating Form	1
Crandall Cognitions Inventory	1
Depression Behaviour Rating Scale	1
Depression Consequences Scale	1
Domains of Satisfaction Scale	2
Dysfunctional Attitude Scale	5
Eysenck Personality Inventory	3
Geriatric Depression Scale	1
Global Assessment Scale	3
Global Illness Rating	1
Global Improvement Scale	1
Global Severity Index	2
Grinker Interview Checklist	1
HRSD	25
Hopkins SCL-90	11
Imaginal Processes Inventory	1
Index of General Affect	1
Internal--External Locus of Control Scale	1
Interpersonal Dependency Scale	1
Interpersonal Events Schedule	1
Inventory for Depressive Symptomatology	1
Inventory of Interpersonal Problems	2
Irrational Beliefs Scale	2
Irritability, Depression Anxiety Questionnaire	1
Kelly's Repartory Grid	1
Loneliness Scale	1
Longitudinal Interval Follow-up Evaluation II	3
LDACL	7
MMPI	21
Modified Personal Orientation Inventory	1
Montgomery–Asberg Depression Scale	2
Mood Perception Inventory	1
Multiple Affect Adjective Checklist	1
NEO Personality Inventory	1
Perceived Social Support Inventory	1
Personal Beliefs Inventory	1
Personal Mood Inventory	1
Personality Assessment Schedule	1
Personality Data Form	1
PES	16
Problems Checklist	1
Problem-solving Inventory	2
Profile of Mood States	1
Psychiatric Outpatients Mood Scale	1
Raskin Eligibility Depression Scale	1
RGSD	5
Rathus Assertiveness Schedule	2
Rosenberg Self-esteem Inventory	2
Rotter Internal–External Locus of Control Scale	1
SADS – Change	4
Self-control Questionnaire	3
Self-control Scale	1
Self-esteem Inventory	2
Self-esteem Reaction Scale	1
Self-esteem Scale	1
Social Adjustment Scale	7
Social Functioning Scale	1
Social Readjustment Rating Scale	1
Social Stress and Support Interview	1
State–Trait Anxiety Inventory	1
Structured Clinical Interview for DSM III R	1
Subjective Probability Questionnaire	1
Symptom Distress Checklist	1
Target Complaints Discomfort Box Scale	1
Tennessee Self-concept	2
Trail-making Test	1
Truax Accurate Empathy Scale	1
Unpleasant Events Schedule	1
Verbaliser–Visualiser Questionnaire	1
Visual Analogue Scale	2
Wolpe–Lazarus Assertiveness Scale	2
Working Alliance Inventory	2
ZSRDS	7

Appendix 10

QRS scores of the included studies

Study number*	Objectives	Sample size	Trial duration	Power calculation	Allocation method	Allocation concealment	Treatment	Subject blinding	Subject source	Diagnostic criteria	Exclusion criteria	Demographic
1	1	0	2	1	2	0	1	0	2	2	1	1
2	1	0	2	0	2	0	1	0	1	0	0	0
3	1	0	0	0	2	0	2	1	1	0	0	0
4	2	0	2	0	2	0	2	1	2	2	2	2
5	2	0	2	1	2	0	2	0	1	2	2	0
6	2	0	2	0	2	0	1	0	1	1	1	2
7	1	0	0	0	0	0	2	0	1	0	0	0
8	1	0	0	0	2	0	2	0	1	0	1	0
9	1	0	0	0	2	0	2	0	1	0	0	0
10	0	0	2	0	2	0	0	1	2	1	2	2
11	1	0	2	0	2	0	1	0	1	2	2	2
12	1	0	2	0	2	0	2	0	2	2	2	2
13	2	2	2	2	2	0	2	0	2	2	1	2
14	2	0	0	0	2	0	2	1	1	0	1	0
15	1	0	0	0	1	0	1	0	0	1	0	0
16	1	0	1	0	2	0	1	0	1	1	0	0
17	1	0	2	0	2	0	2	0	1	2	1	2
18	1	0	0	0	2	0	2	0	1	1	1	0
19	1	0	0	0	2	0	2	0	1	0	0	0
20	2	0	2	1	2	0	2	0	1	2	1	2
21	1	0	1	0	2	0	1	0	1	0	0	1
22	2	0	1	0	2	0	2	0	1	1	1	0
23	2	1	2	0	2	0	2	0	2	2	1	2
24	1	0	1	0	0	0	2	0	1	0	0	0
25	2	2	2	2	2	0	1	0	2	2	2	2
26	1	0	1	0	0	0	2	0	1	2	2	2
27	0	0	1	0	1	0	2	0	1	0	1	1
28	2	0	2	0	1	0	2	0	1	2	1	1
29	1	0	1	0	2	0	2	0	1	1	2	0
30	2	0	0	0	2	0	2	0	1	0	0	1
31	1	0	1	0	2	0	2	0	1	2	2	2

Study number*	Objectives	Sample size	Trial duration	Power calculation	Allocation method	Allocation concealment	Treatment	Subject blinding	Subject source	Diagnostic criteria	Exclusion criteria	Demographic
32	1	0	0	0	2	0	0	0	1	2	0	2
33	1	0	2	0	2	0	2	0	1	2	2	0
34	1	0	2	0	2	0	2	0	1	2	2	0
35	2	0	0	0	2	0	2	0	1	0	2	0
36	1	0	0	0	1	0	2	0	1	1	0	1
37	2	0	0	0	1	0	2	0	1	1	1	0
38	2	0	0	0	1	0	2	0	1	2	1	2
39	2	0	2	0	1	0	2	0	1	2	1	1
40	1	0	2	0	2	0	1	0	2	2	0	1
41	1	0	0	0	1	0	2	0	0	2	1	2
42	1	0	0	0	2	0	1	0	2	2	1	0
42	1	0	1	0	2	1	2	0	1	0	1	2
44	1	1	2	2	2	0	2	0	2	2	2	2
45	1	0	1	0	2	2	1	0	2	1	2	1
46	1	0	2	0	2	0	1	0	2	2	2	1
47	1	0	2	0	2	0	0	0	1	1	0	0
48	1	0	1	0	2	0	2	0	1	2	0	2
49	2	0	2	2	2	1	2	0	1	2	1	1
50	2	0	0	0	2	0	2	0	1	1	1	1
51	0	0	0	0	0	0	1	0	1	1	0	0
52	0	0	0	0	0	0	1	0	1	1	0	0
53	0	0	0	0	2	0	2	0	1	1	0	0
54	2	0	1	0	2	0	2	0	2	2	0	2
55	2	0	2	0	2	0	2	0	1	0	0	1
56	2	0	0	0	0	0	2	0	1	0	1	0
57	1	0	0	0	2	0	2	0	1	0	0	0
58	1	0	1	0	2	0	1	0	1	2	1	2
59	1	0	0	0	2	0	0	0	0	2	1	0
60	1	0	2	0	2	0	2	0	1	1	1	1
61	1	0	1	0	2	0	2	0	1	2	1	1
62	1	0	1	0	2	0	2	0	1	1	1	1
63	2	0	1	0	1	0	2	0	1	1	1	2

Study number*	Assessor blinding	Treatment compliance	Side-effects	Withdrawals	Measures	Comparability	ITT/ completers	Data presentation	Statistical analysis	Conclusions	Declaration of interest	Total
1	0	0	0	1	0	1	0	0	0	0	0	15
2	0	0	0	0	1	1	0	0	0	0	2	11
3	0	0	2	0	1	0	0	1	1	1	0	13
4	1	2	1	2	2	2	1	2	2	1	2	35
5	1	1	0	1	2	2	1	1	2	1	2	28
6	1	2	0	1	2	2	0	1	1	1	2	25
7	0	0	0	1	1	1	0	1	0	1	0	9
8	0	0	0	0	1	2	0	1	1	1	0	13
9	0	0	0	0	2	0	0	1	1	1	0	11
10	1	0	0	2	2	0	1	0	1	0	2	21
11	1	1	0	0	2	2	1	0	1	1	2	24
12	2	2	0	1	1	1	1	0	0	1	2	26
13	2	0	0	1	2	2	1	1	1	1	2	32
14	0	0	0	2	2	0	0	1	1	1	0	16
15	0	0	0	0	1	0	0	0	1	1	0	7
16	0	0	0	0	2	2	0	0	1	1	0	13
17	0	0	0	1	2	2	0	0	2	2	2	24
18	0	0	0	0	0	1	0	0	1	1	0	11
19	0	0	0	0	1	0	0	0	1	1	0	9
20	0	1	0	2	2	2	0	1	2	2	2	29
21	0	2	0	2	2	2	0	0	1	1	0	17
22	0	0	0	0	2	2	0	2	2	2	0	20
23	0	0	0	1	2	2	1	2	2	2	2	30
24	0	0	0	1	1	0	0	1	0	1	0	9
25	1	0	0	0	2	2	0	0	2	2	2	30
26	2	0	0	1	2	2	0	2	2	2	2	26
27	0	0	1	1	2	1	0	0	1	1	0	14
28	1	1	0	2	2	2	0	1	0	0	2	23
29	0	0	0	1	2	2	0	2	1	1	0	19
30	0	0	0	0	2	2	0	0	1	1	0	13
31	0	2	0	2	1	1	1	2	1	2	2	27

Study number*	Assessor blinding	Treatment compliance	Side-effects	Withdrawals	Measures	Comparability	ITT/ completers	Data presentation	Statistical analysis	Conclusions	Declaration of interest	Total
32	0	0	0	0	2	2	0	0	1	0	0	13
33	0	0	0	2	2	2	0	2	2	1	2	25
34	2	0	0	1	2	2	0	1	1	0	0	21
35	0	1	0	1	2	1	0	1	1	1	0	17
36	0	1	0	1	2	1	0	0	0	0	0	12
37	0	0	0	1	1	1	0	0	1	0	2	14
38	2	0	0	1	2	1	0	2	1	1	2	23
39	2	0	0	1	2	0	0	2	2	2	2	25
40	1	2	0	1	2	1	2	1	1	1	2	24
41	0	2	0	1	0	2	1	0	1	1	2	19
42	0	1	0	1	1	0	0	0	1	0	0	13
43	0	2	0	1	2	2	0	2	2	2	0	24
44	2	0	0	1	1	1	2	1	2	1	0	29
45	1	2	0	1	1	0	2	0	1	0	2	23
46	0	0	0	1	2	1	0	1	1	1	2	22
47	0	2	0	1	1	0	2	2	1	1	0	16
48	1	0	0	2	0	1	2	2	2	1	2	24
49	0	0	0	1	2	1	0	1	2	1	0	22
50	1	0	0	0	2	1	0	0	2	1	0	17
51	0	0	0	0	1	1	0	1	1	0	2	9
52	0	0	0	2	0	0	0	1	1	0	2	9
53	0	0	0	0	2	1	0	1	1	0	0	11
54	2	2	0	1	2	1	0	0	0	0	2	23
55	2	0	0	0	2	2	0	1	1	1	2	21
56	0	0	0	0	0	1	0	1	1	0	0	9
57	0	0	0	0	0	0	0	1	1	1	0	9
58	0	0	0	2	2	1	0	1	1	1	0	19
59	0	0	0	0	1	1	1	1	1	1	0	12
60	1	0	0	2	2	0	0	1	1	2	0	20
61	0	0	0	1	2	0	0	0	1	2	2	19
62	2	2	0	0	2	0	0	1	1	0	2	20
63	2	2	0	1	2	1	0	1	1	1	0	22

* *See appendix 6 for the references for each study number*

Appendix 11

Post-treatment dropouts, mean differences and mean change data of the included studies

Study number*	Dropouts reported	Consistency of dropout data	Comments on dropout data	Measure	Mean differences reported	Comments on mean differences data	Mean change reported	Comments on mean change data	Arms included in the pooled mean comparison	Arms excluded
1	Partial	Not consistent	No data on original individual arm sizes: two–five dropped out in each arm	BDI	Partial	Only data for the total sample	No	–	NA	NA
2	No	NA	Dropouts not reported	BDI	Partial	Descriptive data only	No	–	NA	NA
3	No	NA	Dropouts not reported	BDI	Yes	Dropouts and missing data not reported	No	–	Positive connotation versus waiting list	Negative connotation
4	Yes	Consistent and detailed	–	BDI	Yes	Consistent: no missing data	No	–	Social skills versus PDT	NA
5	Yes	Consistent	Initial arm sizes could only be inferred from other data	BDI	Yes	Consistent but endpoint analysis only	Yes	Sent by author	CT versus focused expressive therapy	NA
6	Yes	Actual number unclear	Exact numbers of dropouts and exclusions unclear	BDI	Yes	Dropouts and missing data not reported	No	–	Individual CBT versus control	NA
7	Partial	Paucity	Pretreatment dropouts only reported	BDI	Yes	Consistent: no missing data	No	–	Overt reward versus waiting list	Reward, punish, society influence
8	No	Paucity	Dropouts not reported	BDI	Partial	No SDs	No	–	NA	NA
9	No	Paucity	Author responded: had no data on dropouts	BDI	Yes	Dropouts, missing data and slight discrepancy in data	No	–	Reframing versus waiting list	Self-control
10	Yes	Paucity	Insufficient data on individual arm	NA	No	No data available	No	–	NA	NA
11	Yes	Data in text and tables consistent	Overall dropouts only, but individual arm dropouts could be inferred from other data	BDI	Partial	No SDs	No	–	NA	NA
12	Yes	Consistent across papers	High premature termination rate	RGSD	Partial	No SDs	No	–	NA	NA
13	Yes	Consistent: extensively reported	–	BDI	Yes	Two endpoint analyses	No	–	CBT versus IPT	NA
14	Yes	Paucity	Dropouts reported, but no dichotomous data	BDI	Yes	Consistent: no missing data	No	–	Paradoxical interpretation/ directive versus control	Three other combi-nations

continued

Study number*	Dropouts reported	Consistency of dropout data	Comments on dropout data	Measure	Mean differences reported	Comments on mean differences data	Mean change reported	Comments on mean change data	Arms included in the pooled mean comparison	Arms excluded
15	Partial	Paucity	Dropouts reported, but no dichotomous data	BDI	Partial	No SDs	No	–	NA	NA
16	Yes	Not consistent	New dichotomous data not quite consistent with paper	BDI	Partial	No SDs	No	–	NA	NA
17	Yes	Consistent in table and text	–	BDI	Yes	Three cases missing in data sent by author	Yes	–	CBT versus PDT	NA
18	No	Paucity	–	BDI	Partial	No SDs	No	–	NA	NA
19	No	Paucity	–	NA	No	No mean differences data or SDs	No	–	NA	NA
20	Yes	Consistent	–	BDI	Yes	Both arms in one category of psychotherapy	No	–	NA	NA
21	Yes	Paucity	Some of waiting-list controls transferred into assertion training group	BDI	Partial	Author responded with mean difference for treatment group only	Yes	Sent by author	Assertion training versus waiting list	NA
22	No: assumed to be 0%	Paucity	–	BDI	Yes	Dropouts and missing data not reported	No	–	CT versus IPT	Waiting list
23	Yes	Slight inconsistency in number randomised	Assumed three patients were never randomised	BDI	Yes	Consistency between text and table	No	–	CBT versus behavioural activation	Automatic thoughts
24	Partial	Paucity	No information on arm sizes	BDI	Partial	Insufficient data	No	–	NA	NA
25	Partial	Paucity	Dropouts reported for CBT only	SCL-20	Yes	Sent by author	Yes	Sent by author	CBT versus treatment as usual	NA
26	Yes	Paucity	Assumed nine patients were never randomised	BDI	Yes	Consistency between paper and author's data	No	–	Comprehensive self-control versus psychotherapy	Self-monitoring and principles
27	Partial	Paucity	No data on arm size/ dropout measure for dichotomous outcomes not clear	BDI	Partial	No SDs	No	–	CT versus PDT; CT versus ST	NA
28	Yes	Paucity	Dropouts replaced in trial but still counted as dropouts in the meta-analysis	BDI	Partial	No SDs	No	–	BT versus PDT; BT versus ST	NA
29	Yes	Consistent and clear	Not explicit, but could be inferred	BDI	Yes	Missing data not reported	No	–	CBT versus ST; CT versus BT	Client-centred

continued

Study number*	Dropouts reported	Consistency of dropout data	Comments on dropout data	Measure	Mean differences reported	Comments on mean differences data	Mean change reported	Comments on mean change data	Arms included in the pooled mean comparison	Arms excluded
30	Yes	Consistent	Specific mention of dropouts in text and figures provided	BDI	Partial	Reported by subgroup then combined	No	–	BT versus waiting list	Client-centred
31	Yes	Consistent and clear	–	BDI	No	No data available	Yes	Consistent	CBT versus relaxation	NA
32	No	Paucity	Dropouts not reported	NA	No	No data available	No	–	NA	NA
33	Yes	Consistent in all sections of text	–	BDI	Yes	Consistent with no missing data	No	–	Problem-solving versus waiting list	Problem-focused
34	Yes	Consistent in all sections of text	–	BDI	Yes	Consistent with no missing data	No	–	Problem-solving versus waiting list	NA
35	Yes	Paucity	Dichotomous data reported in % only	BDI	Yes	Consistent with no missing data	Partial	No SDs available	CT versus waiting list	NA
36	Yes	Consistent	Dropouts during treatment but data on all subjects	ZSRDS	Yes	Consistent with no missing data	Yes	Consistent	BT versus ST	NA
37	Yes	Consistent	–	BDI	Partial	No SDs, but a graph	Partial	No SDs	NA	NA
38	Yes	Minor inconsistency	Minor inconsistency in footnote of table	BDI	Yes	Missing data not reported	No	–	Self-control therapy versus waiting list	NA
39	Yes	Paucity	A total of 34 dropouts, but no data for individual arms	BDI	Yes	Dropouts and missing data not reported	No	–	CT versus BT	Combination
40	No	NA	Dropouts not reported, but missing data	BDI	Yes	Consistent with no missing data	No	–	CT versus waiting list	Group CT
41	Yes	Yes	–	BDI	Yes	Dropout data not reported	No	–	Individual versus group	NA
42	Yes	Consistent	–	BDI	Partial	No SDs	No	–	NA	NA
43	Yes	Consistent	–	BDI	Yes	Dropouts and missing data not reported	No	–	CBT versus waiting list	Small versus large group CBT
44	Partial	Paucity	Dropouts reported in treatment group only, but not used	HRSD	Yes	Consistent	Yes	Consistent	IPT versus treatment as usual	NA
45	Yes	Inconsistency between reported and new data	Dropouts reported but others unaccounted for. Available data used	HRSD	Yes	Inconsistency between reported and new data	Yes	Consistent	CT versus treatment as usual	ST
46	Yes	Consistent	–	BDI	Yes	Consistent	Yes	Consistent	CT versus treatment as usual	NA
47	Yes	Consistent in text and table	Dropouts clearly defined/ reported	BDI	Yes	Consistent with no missing data	No	–	Individual CT versus group CT	NA

continued

Study number*	Dropouts reported	Consistency of dropout data	Comments on dropout data	Measure	Mean differences reported	Comments on mean differences data	Mean change reported	Comments on mean change data	Arms included in the pooled mean comparison	Arms excluded
48	Yes	Consistent	No dropouts	BDI	Yes	Consistent with no missing data	No	–	CBT versus waiting list	Computer CBT
49	Yes	Inconsistency between reported and new data	18 dropouts replaced: no further information. Available data used	BDI	Partial	Insufficient data	Yes	Missing data not available	CBT 8 versus PDT/IPT 8; CBT 16 versus PDT/IPT 16	NA
50	No	NA	Dropouts not reported	BDI	Partial	Dropouts and missing data not reported	No	Consistent	CT versus treatment as usual	NA
51	No	NA	Dropouts not reported	MMPI-D	Yes	Missing data not reported	No	–	NA	NA
52	Yes	Consistent	Data from 'expectancy' arms used only	MMPI-D	Yes	Missing data not reported	No	–	Problem-solving versus waiting list	NA
53	No	NA	Dropouts not reported: only waiting-list arm size reported	BDI	Yes	Two of waiting-list group violated protocol but included in analysis	No	–	Problem-solving expectancy versus interest support	NA
54	Yes	Consistent	Dropouts clearly reported	BDI	Partial	No SDs. No missing data, but insufficient data	No	–	CBT versus waiting list	CBT, waiting-list group
55	No	NA	Dropouts not reported, 66 depressed patients taken from sample of 83	BDI	Yes	No missing data	No	–	NA	NA
56	Yes	Replacements used	Dropouts reported but replaced	LDACL	Partial	No SDs; missing data not reported	No	–	NA	NA
57	Partial	Inconsistencies between text and new data	No further data – eight patients unaccounted for, therefore, not used	MMPI	Yes	Missing data not reported	No	–	Gestalt versus minimum contact	Top/underdog, neutral effect
58	Yes	Consistent	High dropout rate from CBT arm	BDI	Yes	Discrepancy between text/new data	Yes	–	CBT versus waiting list	NA
59	No	NA	Dropouts not reported	BDI	Yes	No missing data	No	–	Individual CT versus group CT	NA
60	Yes	Replacements used	Dropouts reported, but replaced. Available data used	BDI	Yes	Unclear whether data included replacements	No	–	CT versus waiting list; CT versus BT	Waiting-list group
61	Partial	Dropouts following randomisation unclear	Only dropouts during treatment (not post-randomisation) reported	MMPI	Partial	No SDs; missing data partially explained	No	–	NA	NA
62	Partial	Six patients dropped out overall	Dropouts per arm and initial arm sizes not reported	BDI	Yes	Missing data not reported	No	–	NA	NA
63	Yes	Consistent	Dropouts clearly reported	BDI	Yes	No missing data	No	–	Individual CT versus group CT	NA

* See appendix 6 for the references for the study *numbers; NA, not applicable*

Appendix 12

Follow-up mean differences and mean change data of the included studies

Study number*	Measure	Times of follow-up for mean differences	Follow-up mean differences reported	Follow-up mean change reported	Comments on mean comparison data	Arms included in pooled mean	Arms excluded
1	BDI	3 months and 1 year	Partial	No	Cells combined to compare with other dataset	Four cells combined	NA
2	BDI	1, 2 and 7 months	Yes	No	Comparison between same intervention	Immediate BT versus delayed BT	NA
3	BDI	3 weeks	Yes	No	Missing data not reported	Positive connotation versus control	Negative connotation
4	BDI	NA	NA	NA	Not eligible for follow-up	NA	NA
5	BDI	3 months	Yes	Yes	Based on same arm sizes as post-treatment (endpoint)	CT versus focused expressive	NA
6	BDI	1 and 6 months	Yes	No	Missing data not reported	Individual CBT versus waiting list	NA
7	BDI	2 weeks and 2 years (partial)	Yes	No	–	Overt reward versus waiting list	Reward, punish and societal influence
8	BDI	5 weeks	Partial	No	No SDs	NA	NA
9	BDI	2 weeks	Yes	No	Missing data not reported	Reframing versus control	Self-control
10	NA	NA	NA	NA	Not eligible for follow-up	NA	NA
11	BDI	3, 6 and 9 months	No	No	Insufficient data	NA	NA
12	RGSD	1 year	Partial	No	Insufficient data	IPT versus ST	NA
13	NA	6, 12 and 18 months	No	No	–	NA	NA
14	BDI	4 weeks	Yes	No	–	Paradoxical interpretation/ paradoxical directive versus control	Three other combinations
15	BDI	6 weeks	Partial	No	No SDs	NA	NA
16	BDI	6 weeks	Partial	No	No SDs	NA	NA
17	BDI	3 months and 1 year	Yes	Yes	New data provided by author	CBT versus PDT	NA
18	BDI	5 weeks	No	No	–	NA	NA
19	NA	NA	NA	NA	No follow-up reported	NA	NA
20	BDI	6 months	Yes	No	Both arms in one category of psychotherapy	NA	NA
21	BDI	8 weeks	No	No	–	NA	NA
22	BDI	4–6 weeks	Yes	No	Missing data not reported	CT versus IPT	Waiting list
23	BDI	6 months	Yes	No	–	CT versus BT	Automatic thoughts
24	BDI	6 months	No	No	–	NA	NA
25	SCL-20	3 and 6 months	Yes	Yes	–	CBT versus treatment as usual	NA
26	BDI	3 months	Yes	No	–	Comprehensive self-control versus PDT	Self-monitoring and principles
27	NA	2 months	Partial	No	No SDs	CT versus PDT	
28	BDI	3, 9, 15, 21 and 27 months	Partial	Partial	No SDs	BT versus PDT	ST
29	BDI	2 months	Yes	No	Missing data not reported	CBT versus ST; CT versus BT	Client-centred

continued

Study number*	Measure	Times of follow-up for mean differences	Follow-up mean differences reported	Follow-up mean change reported	Comments on mean comparison data	Arms included in pooled mean	Arms excluded
30	BDI	6 weeks	Partial	No	No SDs	NA	NA
31	NA	NA	NA	NA	No follow-up reported	CBT versus relaxation training	NA
32	NA	NA	NA	NA	No follow-up reported	NA	NA
33	BDI	6 months	Yes	No	–	Problem-solving versus waiting list	Problem-focused
34	BDI	6 months	No	No	Second arm not eligible for follow-up	NA	NA
35	BDI	4 weeks	Yes	Partial	Paucity, no SDs for mean change	CT versus waiting list	NA
36	NA	NA	NA	NA	No follow-up reported	NA	NA
37	BDI	5–7 weeks	Partial	Partial	Paucity, no SDs	Self control versus assertion	NA
38	NA	NA	NA	NA	No follow-up reported	NA	NA
39	BDI	6 months	Yes	No	–	CT versus BT	CBT
40	BDI	3 and 6 months and 1 year	Yes	No	Only available for group and individual CT	Group versus individual CT	NA
41	BDI	2 months	No	No	No data available	–	–
42	NA	1 month	Partial	No	No SDs or missing data reported	Assertion versus traditional	NA
43	BDI	8 weeks	Yes	No	–	CBT versus waiting list	Small and large group CBT
44	HRSD	8 months	Yes	Yes	–	IPT versus treatment as usual	NA
45	NA	NA	No	No	No follow-up reported	NA	NA
46	BDI	3 and 6 months and 1 year	Yes	Yes	–	CT versus treatment as usual	NA
47	BDI	1, 2, 3 and 6 months	Yes	No	No dropouts for individual CT?	Individual CT versus group CT	NA
48	BDI	2 months	Yes	No	–	Therapist CBT versus waiting list	Computer CBT
49	BDI	3 months and 1 year	Partial	No	No SDs or missing data reported	NA	NA
50	NA	1 month	Partial	No	No SDs or missing data reported	NA	NA
51	NA	NA	NA	NA	No follow-up reported	NA	NA
52	NA	NA	NA	NA	No follow-up reported	NA	NA
53	BDI	5 weeks	Yes	No	Missing data not reported	CT versus BT	CBT, waiting list
54	BDI	3 months	Partial	No	–	CT versus treatment as usual	NA
55	BDI	6 months	Yes	No	No missing data for dichotomous follow-up	Individual versus group CT	NA
56	NA	NA	NA	NA	No follow-up reported	NA	NA
57	MMPI	8 weeks	No	No	Insufficient data	NA	NA
58	NA	NA	NA	NA	No data available for waiting-list group	NA	NA
59	NA	NA	NA	NA	No follow-up reported	NA	NA
60	BDI	5 months	Yes	No	Arm sizes not reported	Insufficient data	Waiting list
61	MMPI	1 and 2 months	Partial	No	No SDs; missing data only partially reported	NA	NA
62	BDI	2 months	Yes	No	All three arms were components of CT	NA	NA
63	BDI	2 months	Yes	No	–	Individual versus group CT	NA

* *See appendix 6 for the references for each study*

Health Technology Assessment Programme

Prioritisation Strategy Group

Members

Chair
Professor Kent Woods
Director,
NHS HTA Programme, &
Professor of Therapeutics
University of Leicester

Professor Bruce Campbell
Consultant General Surgeon
Royal Devon & Exeter Hospital

Professor Shah Ebrahim
Professor of Epidemiology
of Ageing
University of Bristol

Dr John Reynolds
Clinical Director
Acute General Medicine SDU
Oxford Radcliffe Hospital

Dr Ron Zimmern
Director, Public Health
Genetics Unit
Strangeways Research
Laboratories, Cambridge

HTA Commissioning Board

Members

Programme Director
Professor Kent Woods
Director, NHS HTA
Programme, &
Professor of Therapeutics
University of Leicester

Chair
Professor Shah Ebrahim
Professor of Epidemiology
of Ageing
University of Bristol

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

Professor Douglas Altman
Director, ICRF Medical
Statistics Group
University of Oxford

Professor John Bond
Director, Centre for Health
Services Research
University of Newcastle-
upon-Tyne

Ms Christine Clark
Freelance Medical Writer
Bury, Lancs

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

Dr Andrew Farmer
General Practitioner &
NHS R&D
Clinical Scientist
Institute of Health Sciences
University of Oxford

Professor Adrian Grant
Director, Health Services
Research Unit
University of Aberdeen

Dr Alastair Gray
Director, Health Economics
Research Centre
Institute of Health Sciences
University of Oxford

Professor Mark Haggard
Director, MRC Institute
of Hearing Research
University of Nottingham

Professor Jenny Hewison
Senior Lecturer
School of Psychology
University of Leeds

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

Dr Donna Lamping
Head, Health Services
Research Unit
London School of Hygiene
& Tropical Medicine

Professor David Neal
Professor of Surgery
University of Newcastle-
upon-Tyne

Professor Gillian Parker
Nuffield Professor of
Community Care
University of Leicester

Dr Tim Peters
Reader in Medical Statistics
University of Bristol

Professor Martin Severs
Professor in Elderly
Health Care
University of Portsmouth

Dr Sarah Stewart-Brown
Director, Health Services
Research Unit
University of Oxford

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

Dr Gillian Vivian
Consultant in Nuclear
Medicine & Radiology
Royal Cornwall Hospitals Trust
Truro

Professor Graham Watt
Department of
General Practice
University of Glasgow

Dr Jeremy Wyatt
Senior Fellow
Health Knowledge
Management Centre
University College London

continued

continued

Diagnostic Technologies & Screening Panel

Members

Chair
Dr Ron Zimmern
Director, Public Health Genetics Unit
Strangeways Research Laboratories
Cambridge

Dr Philip J Ayres
Consultant in Epidemiology & Public Health
The Leeds Teaching Hospitals NHS Trust

Mrs Stella Burnside
Chief Executive, Altnagelvin Hospitals Health & Social Services Trust
Londonderry
Northern Ireland

Dr Paul O Collinson
Consultant Chemical Pathologist & Senior Lecturer
St George's Hospital, London

Dr Barry Cookson
Director, Laboratory of Hospital Infection
Public Health Laboratory Service, London

Professor Howard Cuckle
Professor of Reproductive Epidemiology
University of Leeds

Dr Carol Dezateux
Senior Lecturer in Paediatric Epidemiology
Institute of Child Health
London

Professor Adrian K Dixon
Professor of Radiology
Addenbrooke's Hospital
Cambridge

Mr Steve Ebdon-Jackson
Head, Diagnostic Imaging & Radiation Protection Team
Department of Health, London

Dr Tom Fahey
Senior Lecturer in General Practice
University of Bristol

Dr Andrew Farmer
General Practitioner & NHS Clinical Scientist
Institute of Health Sciences
University of Oxford

Mrs Gillian Fletcher
Antenatal Teacher & Tutor
National Childbirth Trust
Reigate

Professor Jane Franklyn
Professor of Medicine
University of Birmingham

Dr JA Muir Gray
Joint Director, National Screening Committee
NHS Executive, Oxford

Dr Peter Howlett
Executive Director – Development
Portsmouth Hospitals NHS Trust

Professor Alistair McGuire
Professor of Health Economics
City University, London

Mrs Kathlyn Slack
Professional Support
Diagnostic Imaging & Radiation Protection Team
Department of Health
London

Mr Tony Tester
Chief Officer, South Bedfordshire Community Health Council
Luton

Pharmaceuticals Panel

Members

Chair
Dr John Reynolds
Clinical Director – Acute General Medicine SDU
Oxford Radcliffe Hospital

Dr Felicity J Gabbay
Managing Director, Transcrip Ltd
Milford-on-Sea, Hants

Mr Peter Golightly
Director, Trent Drug Information Services
Leicester Royal Infirmary

Dr Alastair Gray
Director, Health Economics Research Centre
Institute of Health Sciences
University of Oxford

Mrs Jeannette Howe
Senior Principal Pharmacist
Department of Health, London

Dr Andrew Mortimore
Consultant in Public Health Medicine
Southampton & South West Hants Health Authority

Mr Nigel Offen
Head of Clinical Quality
NHS Executive – Eastern
Milton Keynes

Professor Robert Peveler
Professor of Liaison Psychiatry
Royal South Hants Hospital
Southampton

Mrs Marianne Rigge
Director, College of Health
London

Dr Frances Rotblat
Manager, Biotechnology Group
Medicines Control Agency
London

Mr Bill Sang
Chief Executive
Salford Royal Hospitals NHS Trust

Dr Eamonn Sheridan
Consultant in Clinical Genetics
St James's University Hospital
Leeds

Mrs Katrina Simister
New Products Manager
National Prescribing Centre
Liverpool

Dr Ross Taylor
Senior Lecturer
Department of General Practice & Primary Care
University of Aberdeen

Dr Richard Tiner
Medical Director
Association of the British Pharmaceutical Industry
London

Professor Jenifer Wilson-Barnett
Head, Florence Nightingale Division of Nursing & Midwifery
King's College, London

Mr David J Wright
Chief Executive
International Glaucoma Association, London

Therapeutic Procedures Panel

Members

Chair
Professor Bruce Campbell
Consultant General Surgeon
Royal Devon & Exeter Hospital

Professor John Bond
Professor of
Health Services Research
University of Newcastle-upon-Tyne

Ms Judith Brodie
Head of Cancer
Support Service
Cancer BACUP, London

Ms Tracy Bury
Head of Research
& Development
Chartered Society of
Physiotherapy, London

Mr Michael Clancy
Consultant in A&E Medicine
Southampton General Hospital

Professor Collette Clifford
Professor of Nursing
University of Birmingham

Dr Katherine Darton
Information Unit
MIND – The Mental Health
Charity, London

Mr John Dunning
Consultant Cardiothoracic
Surgeon
Papworth Hospital NHS Trust
Cambridge

Mr Jonothan Earnshaw
Consultant Vascular Surgeon
Gloucestershire Royal Hospital

Professor David Field
Professor of Neonatal Medicine
The Leicester Royal Infirmary
NHS Trust

Professor FD Richard Hobbs
Professor of Primary Care
& General Practice
University of Birmingham

Mr Richard Johanson
Consultant & Senior Lecturer
North Staffordshire Infirmary
NHS Trust, Stoke-on-Trent

Dr Duncan Keeley
General Practitioner
Thame, Oxon

Dr Phillip Leech
Principal Medical Officer
Department of Health, London

Professor James Lindesay
Professor of Psychiatry for
the Elderly
University of Leicester

Professor Rajan Madhok
Director of Health Policy
& Public Health
East Riding & Hull
Health Authority

Dr Mike McGovern
Branch Head
Department of Health
London

Dr John C Pounsford
Consultant Physician
Frenchay Healthcare Trust
Bristol

Dr Mark Sculpher
Senior Research Fellow in
Health Economics
University of York

Dr Ken Stein
Consultant in Public
Health Medicine
North & East Devon
Health Authority, Exeter

Expert Advisory Network

Members

Professor John Brazier
Director of Health Economics
University of Sheffield

Mr Shaun Brogan
Chief Executive, Ridgeway
Primary Care Group
Aylesbury, Bucks

Mr John A Cairns
Director, Health Economics
Research Unit
University of Aberdeen

Dr Nicky Cullum
Reader in Health Studies
University of York

Professor Pam Enderby
Chair of Community
Rehabilitation
University of Sheffield

Mr Leonard R Fenwick
Chief Executive
Freeman Hospital
Newcastle-upon-Tyne

Ms Grace Gibbs
Deputy Chief Executive
West Middlesex
University Hospital

Dr Neville Goodman
Consultant Anaesthetist
Southmead Hospital, Bristol

Professor Robert E Hawkins
CRC Professor & Director of
Medical Oncology
Christie Hospital NHS Trust
Manchester

Professor Allen Hutchinson
Director of Public Health &
Deputy Dean, ScHARR
University of Sheffield

Professor David Mant
Professor of General Practice
Institute of Health Sciences
University of Oxford

Professor Alexander Markham
Director
Molecular Medicine Unit
St James's University Hospital
Leeds

Dr Chris McCall
General Practitioner
Corfe Mullen, Dorset

Dr Peter Moore
Freelance Science Writer
Ashtead, Surrey

Dr Sue Moss
Associate Director, Cancer
Screening Evaluation Unit
Institute of Cancer Research
Sutton, Surrey

Mrs Julietta Patnick
National Coordinator
NHS Cancer Screening
Programmes, Sheffield

Professor Jennie Popay
Professor of Sociology &
Community Health
University of Salford

Professor Chris Price
Professor of
Clinical Biochemistry
St Bartholomew's & The
Royal London School of
Medicine & Dentistry

Mr Simon Robbins
Chief Executive
Camden & Islington
Health Authority, London

Dr William Rosenberg
Senior Lecturer &
Consultant in Medicine
University of Southampton

Dr Sarah Stewart-Brown
Director, Health Services
Research Unit
University of Oxford

Dr Gillian Vivian
Consultant in Nuclear
Medicine & Radiology
Royal Cornwall Hospitals Trust
Truro

Mrs Joan Webster
Former Chair
Southern Derbyshire
Community Health Council
Nottingham

Feedback

The HTA Programme and the authors would like to know your views about this report.

The Correspondence Page on the HTA website (http://www.ncchta.org) 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.

Copies of this report can be obtained from:

The National Coordinating Centre for Health Technology Assessment,
Mailpoint 728, Boldrewood,
University of Southampton,
Southampton, SO16 7PX, UK.
Fax: +44 (0) 23 8059 5639 Email: hta@soton.ac.uk
http://www.ncchta.org

ISSN 1366-5278