Who should get low-intensity treatments for depression? An individual patient data meta-analysis

Peter Bower, Professor of Health Services Research, NIHR School for Primary Care Research, Manchester Academic Health Science Centre, University of Manchester, M13 9PL

Evangelos Kontopantelis, Research Fellow, NIHR School for Primary Care Research, Manchester Academic Health Science Centre, University of Manchester, M13 9PL

Alex Sutton, Professor of Medical Statistics, Department of Health Sciences, University of Leicester, Adrian Building, Leicester, LE1 7RH

Tony Kendrick, Professor of Primary Care and Dean, Hull York Medical School, University of York, YO10 5DD

David Richards, Professor of Mental Health Services Research, Washington Singer Laboratories, Psychology, University of Exeter, Exeter, EX4 4QG

Simon Gilbody, Professor of Mental Health Services Research, Department of Health Sciences, University of York, YO10 5DD

Sarah Knowles, Research Associate, NIHR School for Primary Care Research, Manchester Academic Health Science Centre, University of Manchester, M13 9PL

Pim Cuijpers, Professor of Clinical Psychology, Department of Clinical Psychology and EMGO Institute for Health and Care Research, VU University and VU University Medical Center Amsterdam, 1081 BT Amsterdam

Gerhard Andersson, Professor of Clinical Psychology, Department of Behavioural Sciences and Learning, Linköping University, Linköping, Sweden and Karolinska Institute, Stockholm, Sweden

Helen Christensen, Professor, Centre for Mental Health Research, Australian National University, Canberra, Australia

Björn Meyer, Research Department, GAIA AG, Gertigstr. 12-14, 22303 Hamburg, Germany

Marcus Huibers, Department of Clinical Psychology, VU University Amsterdam & Department of Clinical Psychological Science, Maastricht University

Filip Smit, Professor of Public Mental Health, Department of Epidemiology and Biostatistics, and EMGO Institute for Health and Care Research, VU University Medical Center, Amsterdam

Annemieke van Straten, Associate Professor in Clinical Psychology, Department of Clinical Psychology, VU University Amsterdam, EMGO Institute for Health and Care Research, The Netherlands

Lisanne Warmerdam, Research Fellow, Department of Clinical Psychology, Vrije Universiteit Amsterdam, Netherlands

Michael Barkham, Professor of Clinical Psychology, Centre for Psychological Services Research, Department of Psychology, The University of Sheffield, S10 2TP

Linda Bilich, Research Fellow, University of Wollongong, New South Wales, Australia

Karina Lovell, Professor of Mental Health, School of Nursing, Midwifery and Social Work, University Place, University of Manchester, Oxford Road M13 9PL

Emily Tung-Hsueh Liu, Assistant Professor of Clinical Psychology, College of Medicine, Fu-Jen Catholic University, Taiwan

Corresponding author: Dr Peter Bower, NIHR School for Primary Care Research, Manchester Academic Health Sciences Centre, University of Manchester, M13 9PL, peter.bower@manchester.ac.uk, telo: 44 161 275 7638

Details of contributors: The original idea for the research was developed by PB, SG, DR and TK. The IPD database was developed by PB and the analysis conducted by EK with support from AS. SK conducted quality assessments and other data extraction. PC, GA, HC, BM, MH, FS, AvS, LW, MB, LB, KL and ETL all supplied data and assisted with queries. PB and EK wrote the paper and all authors commented on drafts. PB is the guarantor.

Ethical approval: Not required for anonymised secondary data analysis

Funding: The study was funded as part of the United Kingdom National Institute of Health Research (NIHR) School for Primary Care Research. The research team were independent from the funding agency. The views expressed in this publication are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health.

Competing interests: BM is currently a full-time employee of GAIA AG, Hamburg, Germany, a company that owns and developed one of the 'low intensity' interventions considered in this paper. PB is a paid scientific consultant to the British Association of Counselling and Psychotherapy. All other authors declare no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years, no other relationships or activities that could appear to have influenced the submitted work .

Abstract

Objectives – To better manage the high prevalence of depression in the community, many services seek to provide simple, accessible and effective treatments (so called 'low intensity' interventions) to the majority of depressed patients in the first instance, with more intensive and costly treatments retained for patients who do not benefit from 'low intensity' interventions. However, despite large numbers of patients receiving such treatments, there is a lack of evidence to assist clinical decision-making about which patients should receive 'low intensity' interventions. We assessed whether initial severity of depression was a key determinant of the benefit that depressed patients derive from 'low-intensity' interventions

Design – Individual patient data meta-analysis, using 16 datasets comparing 'low intensity' interventions with usual care

Setting – Primary care and community settings

Participants – 2470 patients with depression

Interventions – 'Low intensity' interventions for depression, such as guided self-help using written materials and limited professional support, and internet delivered interventions

Main outcome measures – Depression outcomes

Results – Although referred for 'low intensity' interventions, baseline depression severity highlighted that many patients had significant symptoms. We found a significant interaction between baseline severity and treatment effect, suggesting that patients who are more severely depressed at baseline demonstrate larger treatment effects than those who are less severely depressed. However, the magnitude of the interaction was small and may not be clinically significant.

Conclusions - The data suggests that patients with more severe depression at baseline still show at least as good clinical benefit from 'low intensity' interventions as less severely ill patients and could be offered these treatments as part of a stepped care model.

Introduction

Depression is a major cause of disability among populations worldwide[1] and effective management of this debilitating disorder is a key challenge for health care systems. In response, NHS clinical guidelines for depression recommend a 'stepped care' approach.[2] In stepped care, a significant proportion of patients are treated with 'low-intensity' psychological interventions[3], which are generally based on cognitive behaviour therapy (CBT) and delivered via written materials or information technology with limited professional guidance. Evidence suggests 'low-intensity' interventions will provide many patients with significant clinical benefit.[4,5] In stepped care, conventional 'high-intensity' treatments (such as 12-16 sessions of therapist-led CBT) are then only offered to those who fail to respond to initial 'low-intensity' treatments, or those deemed inappropriate for such interventions. 'Low-intensity' interventions are the primary form of care for hundreds of thousands of depressed patients in the UK through the Improving Access to Psychological Therapies (IAPT) scheme.

At present, one of the key variables determining who gets 'low-' and 'high-intensity' psychological therapy is initial severity of depression. However, the precise thresholds used in decision-making vary, and are largely based on epidemiological studies and accumulated clinical experience, rather than high quality evidence of the empirical relationship between initial severity and outcome in 'low intensity' interventions. This is critical, as the proportion of patients with depression receiving 'low-intensity' interventions as a first intervention varies in practice, but is a key driver of the effectiveness of 'stepped care' depression services and patient experience.[6]

Variables which predict response to interventions are described as *moderators* of treatment effect.[7] Despite the existence of a relatively large literature on the effectiveness of 'low-intensity' interventions,[4,8-11] there is relatively little rigorous evidence on the critical clinical question of whether initial severity moderates effectiveness of 'low-intensity' interventions. Study level meta-analyses[11,12] of these relationships lack precision and are vulnerable to ecological bias.[13] Individual studies often report moderators as secondary analyses, but their yield has been limited by their scarcity, selective reporting,[14] inconsistency in the use of

appropriate methods[7,15] and limited power: sample sizes required to achieve power to detect moderators are potentially very high.[16] This has limited the clinical utility of such analyses.

Individual patient data meta-analysis has the potential to overcome these difficulties and place clinical decision-making in stepped care services on a much firmer footing. This form of analysis can overcome sample size and reporting issues, allow the application of standardised analyses across multiple datasets, and can allow more sophisticated modelling of moderator effects, including the inclusion of covariates and imputation of missing data.[17]

We describe an individual patient data meta-analysis of depression severity as a moderator of the effect of 'low-intensity' interventions in depression,[18] to overcome this gap in the published evidence and make a substantive contribution to clinical decision-making about 'what works for whom' in depression.

Methods

Identification of studies

We primarily used published systematic reviews known to the review team as an efficient and effective method to identify trials meeting our inclusion criteria.[4,5,8,10,11,19,19-22] We updated these with additional searches of the Cochrane Library in July 2011 (see 'additional resources' file for search strategy). We also asked authors of studies identified from the published reviews to identify additional published studies and other trials in progress.

Inclusion criteria for studies

Population: We included studies of patients with depression or mixed depression and anxiety, defined on the basis of research or clinical diagnosis, a minimum score on a depression self-report scale, or self-assessment. Studies of patients with anxiety were excluded unless 50% also achieved a depression diagnosis or the mean depression score met common criteria for 'caseness'.

Context: We included patients managed in non-hospital settings (community and primary care), the settings in which 'low-intensity' interventions are most frequently deployed.

Intervention: We defined 'low-intensity' interventions as those designed to help patients manage depressive symptoms, primarily using a health technology such as self-help books, instructional videos or interactive interventions using information technology. These interventions were conducted predominantly independent of professional or paraprofessional contact (defined as 3 hours or less of contact). We excluded self-help groups and any low-intensity treatment delivered as part of a wider intervention such as 'collaborative care'.

Other criteria: To maximise the possibility of data being available, and to ensure that the analyses involved relatively recent low-intensity interventions, we restricted our analysis to trials reported in 2000 or later. We also restricted our analysis to studies with a sample size of more than 50, to ensure that the logistical effort in obtaining, cleaning and organising the data was commensurate with the contribution to the analysis.[17]

The study protocol is available from the first author.

Data preparation and analysis

We sought primary datasets from study authors, with the following core variables: randomised group, baseline depression measures, follow-up depression measures, age and sex. We combined the datasets into a single archive and conducted analyses to ensure that variables were correctly specified and that initial analyses of individual data sets were consistent with published data.

Measure standardisation

Almost all studies either used BDI [24] or CES-D[25] as the main depression outcome. We report scores on these scales for descriptive purposes, converting one trial using the CORE-OM[26] to BDI scores using published algorithms[27] to maximise comparability. For the main

analysis we standardised scores within each study, using study specific means of the follow up scores and the standard deviations of the baseline scores. Patients participating in low-intensity trials may be selected to be appropriate for these interventions, and there may be limits on the severity of patients included in such trials, restricting our ability to test the moderating effects of severity at the higher range. We assessed the severity of patients included in these trials, both in terms of inclusion and exclusion criteria, and the BDI and CES-D scores of patients actually recruited.

Missing data

We assumed data were missing at random and imputed missing age and depression scores at follow up using a multivariate imputation algorithm (*mi impute mvn,* in *Stata* version 11) using Markov Chain Monte Carlo (MCMC). Multiple imputation is currently the most sophisticated approach to deal with missing data and is recommended over single imputation.[28,29] The method generates several datasets, analysing each one separately using the selected model and combines the results. We generated one thousand new datasets with the observed and imputed scores for age and follow-up depression scores from study, treatment group, baseline depression score and sex. Predicted scores were limited to ranges appropriate for each scale. Convergence of the MCMC algorithms was verified with time-series and autocorrelation plots of the worst linear function.[30,31]

Analysis

As individual patient data meta-analyses are vulnerable to publication bias from a number of sources,[23] two authors independently extracted data on populations, interventions, methodological quality (based on assessment of allocation concealment, intention to treat analysis and attrition) and outcome effect sizes for *all* studies identified by the searches, so as to compare the studies where data were available to us with those where it was not. We present descriptive statistics on study characteristics and assessed the potential for publication bias using funnel plots, in line with published recommendations.[23]. We also extracted data on moderator analyses in published studies to allow further comparisons.

There are three methods of analysing moderator effects in meta-analysis: aggregate data analysis through meta-regression; using individual patient data to estimate the treatment-moderator interaction within each study, followed by a standard inverse-variance meta-analysis ('two-step analysis'); and analysis of individual patient data using a mixed model and accounting for clustering of patients within studies ('one-step analysis').[13,17] In certain situations these last two analyses give identical results, although they differ under conditions such as 'covariate heterogeneity' (i.e. the variation in the covariate within each study).[13]

In this study we used the one step analysis, which is the most logistically demanding, but allows for sophisticated modelling of covariates (in this case, age, sex and baseline severity), is least affected by bias and is most efficient in terms of power.[32,33] Appropriate mixed-effects models (with fixed-trial-specific intercepts, a random treatment effect and fixed trial specific effects for baseline) were used to meta-analytically synthesise the patient-level data and estimate the between- and within-study variances, fitting the interaction as a continuous variable.[34] We used *Stata* v12.1 and a restricted maximum likelihood algorithm with the *xtmixed* command.[35] Heterogeneity was assessed using the I^2 statistic.[36] For cluster randomised studies we adjusted appropriately.[37] Where studies involved multiple treatment comparisons with a single control, we treated each comparison separately, and we avoided double counting controls by assigning half the controls at random to each comparison.

We conducted two pre-specified secondary analyses to assess the robustness of the results. We explored whether the overall moderating effects of baseline severity were substantively different at the highest levels of baseline severity (i.e. to test whether there was a non-linear effect at the highest levels of depression severity). We split the data into five equally sized groups on the basis of the initial severity of patients (rather than 2 as specified in the protocol), and assessed the moderating effect of baseline severity in each group.

We also assessed whether the main result was influenced by a single indicator of study quality (a dichotomous measure based on adequacy of allocation concealment), and the types of low intensity interventions: internet versus written forms, and 'guided' (i.e. 'low intensity' interventions with limited support by a health professional) versus 'unguided' forms (used by the patient alone).

Results

The PRISMA diagram is provided in Figure 1. Six potentially eligible studies were excluded because N<50, 5 because they were published before 2000, and 4 were excluded on both criteria. We identified 29 comparisons as being potentially eligible. There was moderate evidence of asymmetry in the funnel plot for these studies (Egger's regression test intercept -2.4 standard error 0.8, p=0.007, Figure 2). We gained access to data from 16 of these comparisons (55%), with data unavailable either because of no response from authors (n=8), clashes with their own planned analyses (n=4) and ethical issues with sharing data (n=1). A small number of individual cases (n=20) were dropped because of missing baseline age or depression scores, leaving 2470 unique cases, with 77% reporting data at first follow up.

Available and unavailable data

Data on study characteristics and design are detailed in the 'additional resources' file. We compared available and unavailable studies on population, intervention, quality and outcome data (see Table 1). Studies were similar in recruitment procedures, although available data were less likely to involve health technologies delivered via information technology, but were more likely to involve support from a health professional. Included studies were generally higher quality and had a larger sample size and reported lower estimates of effect.

Baseline characteristics of patients included in the review

As noted earlier, patients participating in low-intensity trials are selected to be appropriate for these interventions, so we assessed the severity of patients included in these trials. Six (38%) had a maximum ceiling for inclusion. Assessment of baseline means highlighted that many patients had significant symptoms (see Figure 3). For the BDI, a score of 10-16 indicates mild depression, 17-29 indicates moderate depression, and 30+ severe depression, and mean scores were 19-21,[38] 21,[39] 22,[40] 23-24,[41] 23-28,[42] 26,[43], 27,[44] 27-28,[45] and 30.[46] A score of 16+ indicates a probable depressive illness on the CESD, and mean scores ranged from 13 in a trial focussed on subthreshold symptoms,[47] 21-22,[48] 30,[49] and 32.[50]

Other characteristics of the patients are presented in etable x. [add individual patient data on age, sex, treatment history where available]

Is the effect of low-intensity interventions on depression moderated by baseline depression severity?

The overall standardised estimate of the effects of 'low-intensity' interventions was -0.42 (95% CI -0.55 to -0.29, I^2=2.9%, 95% CI 0.5 to 15%). When a term was added to assess the interaction, we found a significant negative interaction between baseline severity and treatment effect (interaction co-efficient -0.1, 95% CI -0.19 to -0.002). This suggests that patients who are more severely depressed at baseline demonstrate larger treatment effects than those who are less severely depressed. However, the magnitude of the interaction is small. As scores had been standardised, the effect represented an additional standardised benefit of 0.1 for an increase in initial severity of 1 standard deviation, which may not be clinically significant. The estimates of the interactions at the level of the individual studies is shown in Figure 4.

Is there a moderating effect of baseline depression severity at higher levels of depression?

The main analysis reported in the previous section showed a small but significant increase in effect of 'low-intensity' interventions in more severe patients. When data were analysed in terms of 5 severity subgroups, we observed a stepwise increase in the effect of 'low-intensity' interventions, from least to most severely ill patients, but there was no statistically significant difference in the effect across the groups. Thus there was no indication that patients at the highest levels of severity showed different effects to the overall trend.

Are the results sensitive to allocation concealment?

The moderating effect of initial depression was larger in patients in studies with adequate concealment of allocation, but the difference was not statistically significant (interaction co-efficient -0.07 95% CI -0.34 to 0.21).

Are the results sensitive to types of 'low intensity' interventions?

Compared with those in studies using written interventions, the moderating effect of initial depression was larger in patients recruited to studies using internet 'low intensity' treatments but the difference was not statistically significant (interaction co-efficient -0.09

95% CI -0.31 to 0.12). Compared with those in studies using guided 'low intensity' interventions, the moderating effect of initial depression was greater in patients using unguided 'low intensity' interventions but the difference was not statistically significant (interaction co-efficient -0.07, 95% CI -0.30 to 0.15).

Discussion

Statement of principal findings

Data from 16 comparisons of 'low-intensity' interventions in depression showed that patients with more severe depression at baseline derive at least as good clinical benefit from 'low intensity' interventions as less severely ill patients. We did not find strong evidence that the main result was dependent on allocation concealment or the types of 'low-intensity' interventions, although these additional secondary analyses lacked precision.

Strengths and weaknesses of the study

Although generally considered as a gold standard, individual patient data meta-analyses are potentially vulnerable to *publication bias* (selective publication of significant results in primary studies), *reviewer selection bias* (selective identification of relevant individual patient data datasets) and *availability bias* (selective access to individual patient datasets once identified). The funnel plot suggested the potential for publication bias in the general literature around 'low intensity' interventions. Reviewer selection bias was reduced by the search methods (using published systematic reviews and a search for recent studies). In terms of availability bias, a recent review found that the proportion of available patients in individual patient data analyses ranged from 66% to 98%.[23] We were unable to access all relevant datasets and our data included only 55% of the total number of patients available. As well as a relatively high level of unavailable data, the trials with available data differed in important ways from the entire literature. The results may not generalise as clearly to patient populations with a formal diagnosis of depression, to computerised 'low intensity' interventions, and to unguided interventions. The diagnosis issue is probably the key limitation, as it relates most clearly to the core research question.

As noted previously, it is possible that patients with very severe depression (and therefore more likely to receive a diagnosis) would not enter these trials, so the analysis is unable to assess their outcomes. However, it should be noted that the 10 trials in the dataset using the BDI included 430 patients (nearly one third of the total) with scores >30 (indicating severe depression), which demonstrates that these samples do not consist of minor cases only. Our secondary analyses did not suggest that the general direction of effects was different in the most severely depressed patients. Figure 3 would suggest that the results are valid with scores between 40 and 50 on the two outcome measures.

Strengths and weaknesses in relation to other studies

There are no comparable analyses in the 'low intensity' intervention literature. Thirteen comparisons in the total dataset included some form of secondary analyses of moderators (see etable 1), although the variables tested and the analytic techniques used varied widely, and not all explored severity. Of those examining initial severity, 4 comparisons suggested similar results in less and more severely ill patients,[48,51,52] one reported a greater benefit in less severely ill patients[46] and the rest reported that more severely ill patients showed greater benefits.[45,49,53] The broad pattern thus confirms the present findings, although issues with the analyses and power of previous studies means that the current analysis has a rigour and precision that a narrative analyses of patterns across individual studies cannot match.

One recent meta-analysis assessed the impact of pre-treatment severity on outcomes in conventional 'high-intensity' psychological therapies for outpatient depression.[12] Meta-regression results showed that mean pre-treatment depression scores did not generally predict intervention effects across all studies. A subset of studies reported within-study analyses, and the data from these suggested that where effects were demonstrated, they concurred with the present analysis in showing that higher initial severity was associated with greater treatment effects.

Meaning of the study: possible explanations and implications for clinicians and policymakers

The results would suggest that it is legitimate to include 'low-intensity' interventions in the first step of a stepped care system and to encourage the majority of patients to use them as

the initial treatment option, even when initial severity of depression is high. Clearly some patients will not find such treatments useful, and it would seem sensible to continue to refer very severe cases to more intense psychological treatment or pharmacological management until further evidence is generated confirming these findings. The current data would suggest that the threshold could be relatively high if patients are willing to engage in 'low intensity' interventions.

There are caveats to that recommendation. It is important to note that we have only modelled the impact of initial severity on the *comparative* effectiveness of 'low intensity' interventions. Even though more severely ill patients show comparable benefit to less severely ill patients, their high initial scores mean that many remain symptomatic and do not meet conventional thresholds for 'recovery'. The second critical aspect of stepped care systems is that all patients are monitored consistently after any treatment to assess progress and ensure that those with residual symptoms receive additional care to enhance the likelihood of long-term recovery.[56] In relation to this, it is possible that immediate provision of high intensity treatments to patients with more severe depression would be more cost effective than use of low intensity treatments first. Secondly, it is possible that initial experience with low intensity treatments (especially if unsuccessful) could act as a barrier to further treatment. Data to explore either of these hypotheses are simply not available at present,

It remains to be seen what other patient factors might need to be taken into account in clinical decision-making. The traditional model of evidence-based practice would suggest that patient needs and preferences are important, but the evidence demonstrating a relationship of preferences to outcome is varied.[54,55] The effects of preferences could in principle be tested in a similar way to the current analysis if baseline data were reported consistently.[55]

Unanswered questions and future research

The current data demonstrates that some of the concerns about examination of moderators in clinical trials (especially those around sample size) can be overcome through collaborative

individual patient data meta-analysis. It is important that the ethical and logistical barriers to such data sharing are removed, and appropriate incentives put in place to encourage such analyses to answer clinically relevant questions in the future.

The present analysis highlights the potential for more effective collaboration around data sharing to enable appropriately powered secondary subgroup analyses, with the potential to allow more effective targeting of treatments to patients and more personalised care. However, it is important to note that there may be far more effective predictors of outcomes than baseline severity, including preferences[55] and other psychological variables relating to attitudes or aptitudes. Fully exploring these issues will require a consistent approach to defining core moderating variable data to be collected at baseline, similar to calls around core outcome measures in trials,[57] to allow development of an evidence base to provide better guidance for patients, health professionals and policy makers about 'what works for whom' in depression.

Reference List

1 Murray C, Lopez A. *The Global Burden of Disease: a comprehensive assessment of mortality and disability from diseases, injuries, and risk factors in 1990 and projected to 2020*. Harvard School of Public Health: Harvard University Press, 1996.

2 National Institute for Health and Clinical Excellence. *Depression: the treatment and management of depression in adults (update)*. http://www.nice.org.uk/nicemedia/pdf/Depression_Update_FULL_GUIDELINE.pdf. National Institute for Health and Clinical Excellence, 2010.

3 Bennett-Levy J, Richards D, Farrand P. Low intensity CBT interventions: a revolution in mental health care. In: J Bennett-Levy, D Richards, P Farrand, et al, eds. *Oxford guide to low intensity CBT interventions*. Oxford: Oxford University Press, 2010;

4 Andrews G, Cuijpers P, Craske M, McEvoy P, Titov N. Computer therapy for the anxiety and depressive disorders is effective, acceptable and practical health care: a meta-analysis. *Plos ONE* 2010;**5**:

5 Cuijpers P, Donker T, Van Straten A, Li J, Andersson G. Is guided self-help as effective as face to face psychotherapy for depression and anxiety disorders? A systematic review and meta-analysis of comparative outcome studies. *Psychol Med* 2010;**40**:1943-1957.

6 Richards D, Bower P, Pagel C, et al. Delivering stepped care for depression: an analysis of implementation in routine practice. *Implementation Science* 2012;**7**:

7 Kraemer H, Wilson G, Fairburn C, Agras W. Mediators and moderators of treatment effects in randomized clinical trials. *Arch Gen Psychiatry* 2002;**59**:877-883.

8 Andersson G, Cuijpers P. Internet-based and other computerized psychological treatments for adult depression: a meta-analysis. *Cognitive Behaviour Therapy* 2010;**38**:196-205.

9 Spek V, Cuijpers P, Nykliceck I, Riper H, Keyzer J, Pop V. Internet-based cognitive behaviour therapy for symptoms of depression and anxiety: a meta-analysis. *Psychol Med* 2007;**37**:1-10.

10 Kaltenthaler E, Parry G, Beverley C. The clinical and cost-effectiveness of computerised cognitive behaviour therapy (CCBT) for anxiety and depression. *Health Technol Assess* 2006;**10 (33)**:

11 Gellatly J, Bower P, Hennessey S, Richards D, Gilbody S, Lovell K. What makes self-help interventions effective in the management of depressive symptoms? Meta-analysis and meta-regression. *Psychol Med* 2007;**37**:1217-1228.

12 Driessen E, Cuijpers P, Hollon S, Dekker J. Does pre-treatment severity moderate the efficacy of psychological treatment of adult outpatient depression? A meta-analysis. *J Consult Clin Psychol* 2010;**78**:668-680.

13 Simmonds M, Higgins J. Covariate heterogeneity in meta-analysis: criteria for deciding between meta-regression and individual patient data. *Stat Med* 2011;**26**:2982-2999.

14 Sun X, Briel M, Busse J, et al. The influence of study characteristics on reporting of subgroup analyses in randomised controlled trials: systematic review. *BMJ* 2012;**342**:d1569

15 Sun X, Briel M, Busse J, et al. Credibility of claims of subgroup effects in randomised controlled trials: systematic review. *BMJ* 2012;**344**:e1553

16 Brookes S, Whitely E, Egger M, Davey Smith G, Mulheran P, Peters T. Subgroup analyses in randomized trials: risks of subgroup-specific analyses; power and sample size for the interaction test. *J Clin Epidemiol* 2011;**57**:229-236.

17 Riley R, Lambert P, Abo-Zaid G. Meta-analysis of individual participant data: rationale, conduct, and reporting. *BMJ* 2010;**340**:c221

18 Stewart L, Clarke M, Cochrane Working Group on meta-analysis using individual patient data. Practical methodology of meta-analyses (overviews) using updated individual patient data. *Stat Med* 1995;**14**:2057-2079.

19 Anderson L, Lewis G, Araya R, et al. Self-help books for depression: how can practitioners and patients make the right choice? *Br J Gen Pract* 2005;**55**:387-392.

20 Bower P, Richards D, Lovell K. The clinical and cost-effectiveness of self-help treatments for anxiety and depressive disorders in primary care: a systematic review. *Br J Gen Pract* 2001;**51**:838-845.

21 Cuijpers P. Bibliotherapy in unipolar depression: a meta-analysis. *J Behav Ther Exp Psychiatry* 1997;**28**:139-147.

22 Kaltenthaler E, Shackley P, Stevens P, Beverley C, Parry G, Chilcott J. A systematic review and economic evaluation of computerised cognitive behaviour therapy for depression and anxiety. *Health Technol Assess* 2002;**6 (22)**:

23 Ahmed I, Sutton A, Riley R. Individual participant data meta-analysis: an assessment of publication bias, selection bias and unavailable data. *BMJ* 2011;**344**:d7762

24 Beck A, Steer R, Carbin M. Psychometric properties of the Beck Depression Inventory: twenty five years of evaluation. *Clin Psychol Rev* 1988;**8**:77-100.

25 Radloff L. The CES-D scale: a self-report depression scale for research in the general population. *Applied Psychological Measurement* 1977;**1**:385-401.

26 Evans C, Mellor-Clark J, Margison F, et al. CORE: clinical outcomes in routine evaluation. *J Ment Health* 2000;**9**:247-255.

27 Leach C, Lucock M, Barkham M, Stiles W, Noble R, Iveson S. Transforming between Beck Depression Inventory and CORE-OM scores in routine clinical practice. *Br J Clin Psychol* 2006;**45**:153-166.

28 Gadbury G, Coffey C, Allison D. Modern statistical methods for handling missing repeated measurements in obesity trial data: Beyond LOCF. *Obesity Reviews* 2003;**4**:175-184.

29 Donders A, van der Heiden G, Stijnen T, Moons K. Review: a gentle introduction to imputation of missing values. *J Clin Epidemiol* 2012;**59**:1087-1091.

30 Schafer J. *Analysis of Incomplete Multivariate Data*. London: Chapman and Hall, 1997.

31 Stata Press. *Multiple-Imputation Reference Manual*. College Station: Stata Press, 2012.

32 Stewart L, Parmar M. Meta-analysis of the literature or of individual patient data: is there a difference? *Lancet* 1993;**341**:418-422.

33 Vickers A, Cronin A, Maschino A, et al. Individual patient data meta-analysis of acupuncture for chronic pain: protocol of the Acupuncture Trialists' Collaboration. *Trials* 2010;**11**:90

34 Whitehead A. *Meta-Analysis Of Controlled Clinical Trials*. New York: John Wiley & Sons, Ltd, 2002.

35 StataCorp LP. Stata Statistical software. [11.1]., 2009: StataCorp LP.

36 Higgins J, Thompson S. Quantifying heterogeneity in a meta-analysis. *Stat Med* 2002;**21**:1539-1558.

37 Sutton A, Kendrick D, Coupland C. Meta-analysis of individual- and aggregate-level data. *Stat Med* 2008;**27**:651-669.

38 Bilich L, Deane F, Phipps A, Barisic M, Gould G. Effectiveness of bibliotherapy self-help for depression with varying levels of telephone helpline support. *Clinical Psychology and Psychotherapy* 2008;**15**:61-74.

39 Andersson G, Bergstrom J, Hollandare F, Carlbring P, Kaldo V, Ekselius L. Internet-based self-help for depression: randomised controlled trial. *The British Journal of Psychiatry* 2005;**187**:456-461.

40 Vernmark K, Lenndin J, Bjärehed J, et al. Internet administered guided self-help versus individualized e-mail therapy: a randomized trial of two versions of CBT for major depression. *Behav Res Ther* 2010;**48**:368-376.

41 Richards A, Barkham M, Cahill J, Richards D, Williams C, Heywood P. PHASE: a randomised, controlled trial of supervised self-help cognitive behavioural therapy in primary care. *Br J Gen Pract* 2003;**53**:764-770.

42 Liu E, Chen W, Li Y, Wang C, Mok T, Huang H. Exploring the efficacy of cognitive bibliotherapy and a potential mechanism of change in the treatment of depressive symptoms among the Chinese: a randomized controlled trial. *Cognit Ther Res* 2009;**33**:449-461.

43 Mead N, MacDonald W, Bower P, Lovell K, Richards D, Bucknall A. The clinical effectiveness of guided self-help versus waiting list control in the management of

anxiety and depression: a randomised controlled trial. *Psychol Med* 2005;**35**:1633-1643.

44 Meyer B, Berger T, Caspar F, Beevers C, Andersson G, Weiss M. Effectiveness of a novel integrative online treatment for depression (Deprexis): randomized controlled trial. *Journal of Medical Internet Research* 2011;**11**:e15

45 de Graaf L, Gerhards S, Amtz A, et al. Clinical effectiveness of online computerised cognitive behavioural therapy without support for depression in pirmary care: randomised trial. *Br J Psychiatry* 2009;**195**:73-80.

46 Lovell K, Bower P, Richards D, et al. Developing guided self-help for depression using the Medical Research Council complex interventions framework: a description of the modelling phase and results of an exploratory randomised controlled trial. *BMC Psychiatry* 2008;**8**:

47 Willemse G, Smit F, Cuijpers P, Tiemens B. Minimal contact psychotherapy for sub-threshold depression in primary care: randomised trial. *Br J Psychiatry* 2004;**185**:416-421.

48 Christensen H, Griffiths K, Jorm A. Delivering interventions for depression by using the internet: randomised controlled trial. *BMJ* 2004;**328**:265

49 Van Straten A, Cuijpers P, Smits N. Effectiveness of a web-based self-help intervention for symptoms of depression, anxiety, and stress: randomized controlled trial. *J Med Internet Res* 2008;**10**:e7

50 Warmerdam L, Van Straten A, Twisk J, Riper H, Cuijpers P. Internet-based treatment for adults with depressive symptoms: randomized controlled trial. *J Med Internet Res* 2008;**10**:e44

51 Proudfoot J, Ryden C, Everitt B, et al. Clinical efficacy of computerised cognitive-behavioural therapy for anxiety and depression in primary care: randomised controlled trial. *Br J Psychiatry* 2004;**185**:46-54.

52 Salkovskis P, Rimes K, Stephenson D, Sacks G, Scott J. A randomized controlled trial of the use of self-help materials in addition to standard general practice treatment of depression compared to standard treatment alone. *Psychol Med* 2006;**36**:325-333.

53 Levin W, Campbell D, McGovern K, et al. A computer-assisted depression intervention in primary care. *Psychol Med* 2010;**41**:1373-1383.

54 King M, Nazareth I, Lampe F, et al. Impact of participant and physician intervention preferences on randomized trials: a systematic review. *JAMA* 2005;**293**:1089-1099.

55 Preference Collaborative Review Group. Patients' preference within randomised trials: systematic review and patient level meta-analysis. *BMJ* 2008;**337**:1864

56 Bower P, Gilbody S. Stepped care in psychological therapies: access, effectiveness and efficiency: narrative literature review. *Br J Psychiatry* 2005;**186**:11-17.

57 Williamson P, Altman D, Blazeby J, Clarke M, Gargon E. Driving up the quality and relevance of research through the use of agreed core outcomes. *J Health Serv Res Policy* 2012;**17**:1-2.

Figure 1 Inclusion of studies in the review

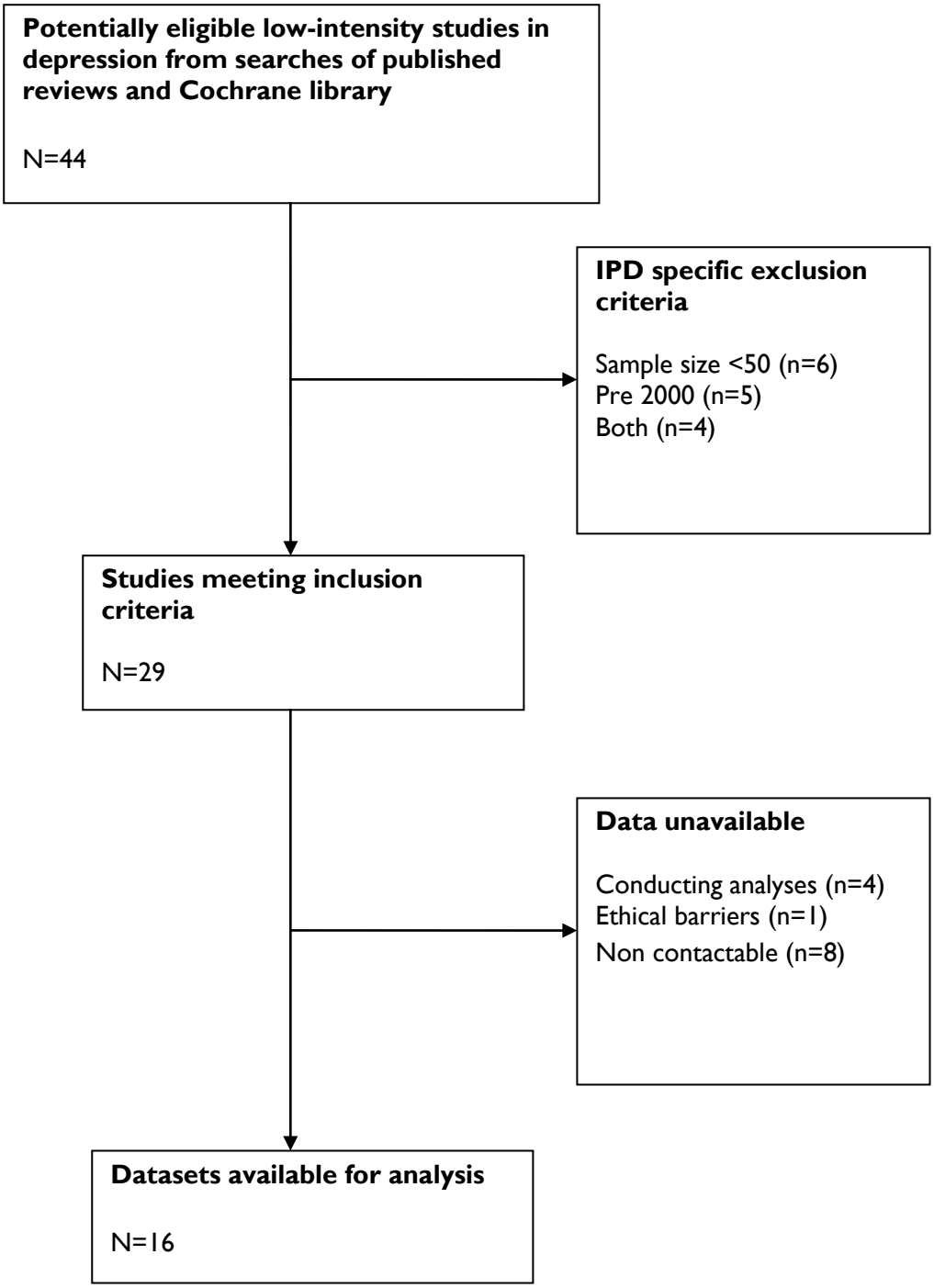
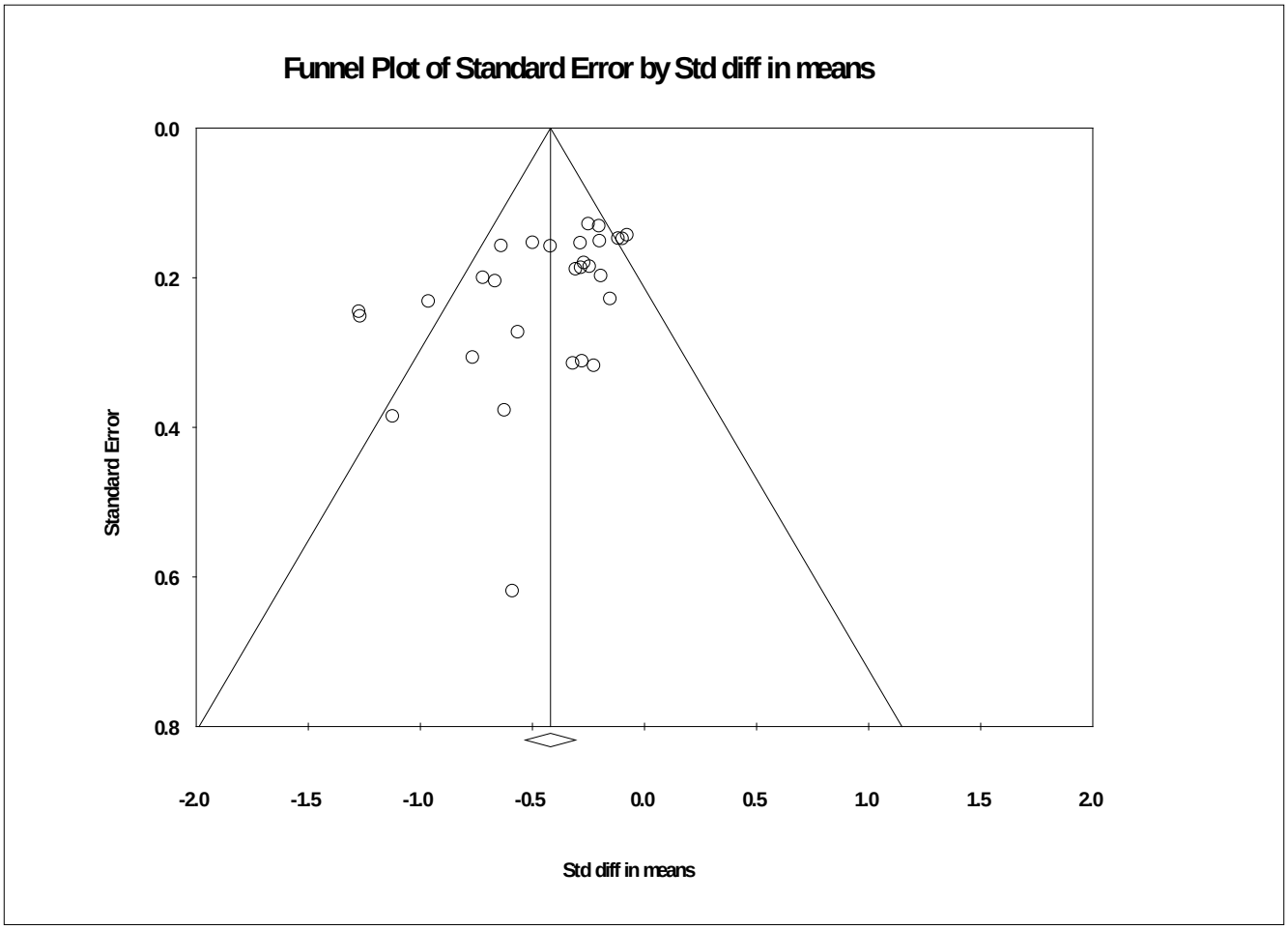


Figure 2 Funnel plot



Egger's regression intercept

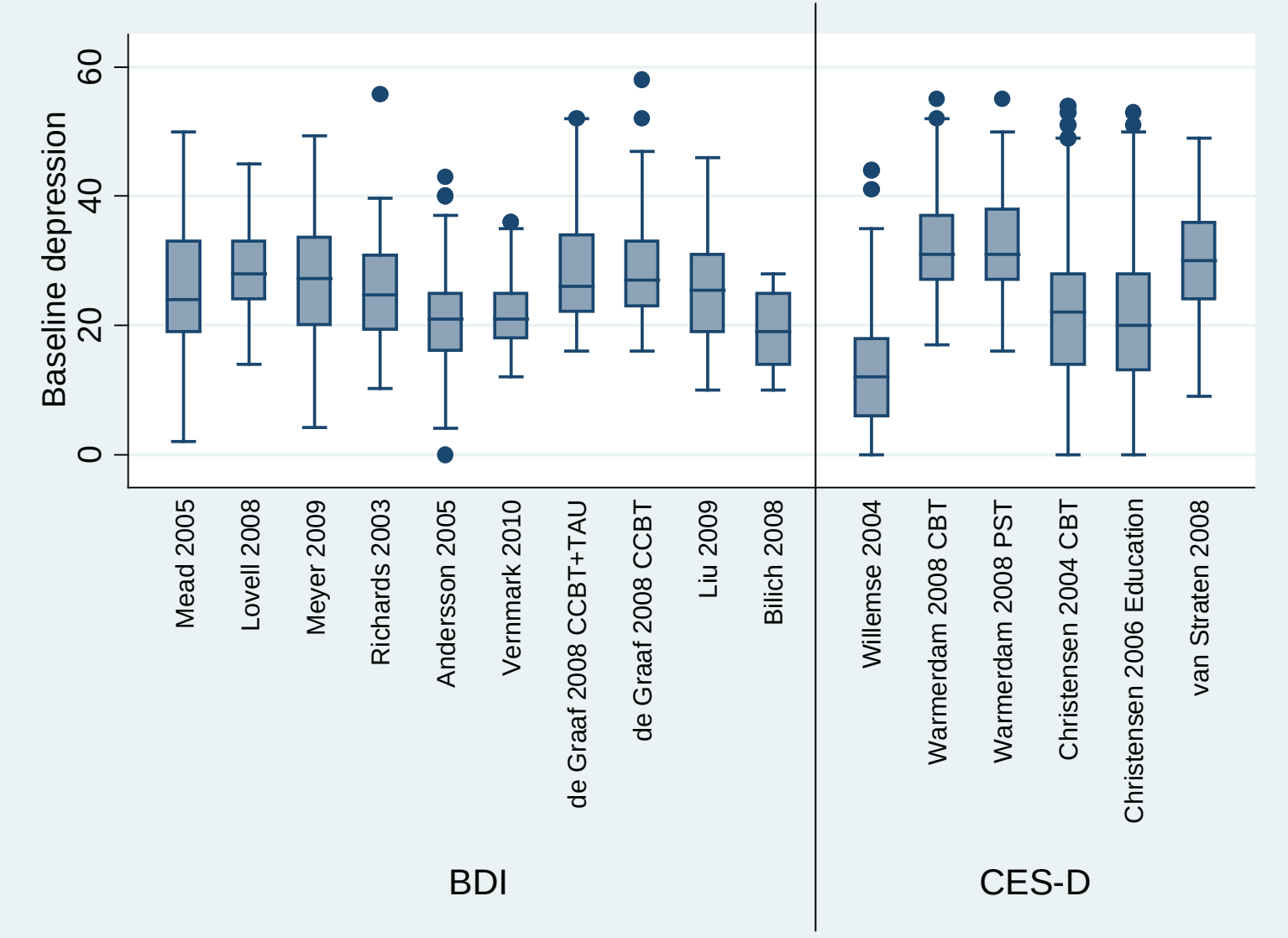
Intercept	-2.39404
Standard error	0.81767
95% lower limit (2-tailed)	-4.07177
95% upper limit (2-tailed)	-0.71632
t-value	2.92788
df	27.00000
P-value (1-tailed)	0.00343
P-value (2-tailed)	0.00685

Table 1 Comparison of available and unavailable studies

Factor	Unavailable (n=13)	Available (n=16)
Recruitment via screening (versus referral)	77%	81%
Depression diagnosis confirmed	46%	13%
Computerised delivery (versus bibliotherapy)	92%	63%
Guided minimal intervention	46%	75%
Mean quality (0-3)[1]	1.9	2.3
Mean baseline N	157	180
Pooled effect size (95% CI)	-0.47 (xx to xx)	-0.39 (xx to xx)

[1] Scored 0 to 3 on the basis of adequacy of concealment of allocation, reporting of intention to treat analysis, and >20% attrition

Figure 3 Baseline severity data of studies included in the review



BDI scores range from 0-63, CESD scores from 0-60. Length of the box represents interquartile range, extent of line represents minimum and maximum scores, other points are outliers

Figure 4 Forest plot of interactions between baseline severity and allocation