Influence of initial severity of depression on effectiveness of low intensity interventions: meta-analysis of individual patient data

Peter Bower, Evangelos Kontopantelis, Alex Sutton, Tony Kendrick, David A. Richards, Simon Gilbody, Sarah Knowles, Pim Cuijpers, Gerhard Andersson, Helen Christensen, Bjoern Meyer, Marcus Huibers, Filip Smit, Annemieke van Straten, Lisanne Warmerdam, Michael Barkham, Linda Bilich, Karina Lovell and Emily Tung-Hsueh Liu

Linköping University Post Print

N.B.: When citing this work, cite the original article.

RESEARCH

Influence of initial severity of depression on effectiveness of low intensity interventions: meta-analysis of individual patient data

OPEN ACCESS

Peter Bower *professor of health services research*[1], Evangelos Kontopantelis *research fellow*[1], Alex Sutton *professor of medical statistics*[2], Tony Kendrick *professor of primary care and dean*[3], David A Richards *professor of mental health services research*[4], Simon Gilbody *professor of psychological medicine & health services research*[5], Sarah Knowles *research fellow*[1], Pim Cuijpers *professor of clinical psychology*[6], Gerhard Andersson *professor of clinical psychology*[7], Helen Christensen *professor and executive director*[8], Björn Meyer *research director and honorary research fellow*[9], Marcus Huibers *professor of psychotherapy*[10], Filip Smit *professor of public mental health*[11], Annemieke van Straten *professor in clinical psychology*[6], Lisanne Warmerdam *research fellow*[6], Michael Barkham *professor of clinical psychology*[12], Linda Bilich *research fellow*[13], Karina Lovell *professor of mental health*[14], Emily Tung-Hsueh Liu *associate professor of clinical psychology*[15]

[1]NIHR School for Primary Care Research, Manchester Academic Health Science Centre, University of Manchester, Manchester M13 9PL, UK; [2]Department of Health Sciences, University of Leicester, Leicester, UK; [3]Hull York Medical School, University of York, York, UK; [4]Sir Henry Wellcome Building, University of Exeter Medical School, University of Exeter, Exeter, UK; [5]Department of Health Sciences, University of York & Hull York Medical School (HYMS); [6]Department of Clinical Psychology and EMGO Institute for Health and Care Research, VU University and VU University Medical Center Amsterdam, Amsterdam, Netherlands; [7]Department of Behavioural Sciences and Learning, Linköping University, Linköping, Sweden, and Karolinska Institute, Stockholm, Sweden; [8]Black Dog Institute, University of New South Wales, Randwick NSW, Australia; [9]Research Department, GAIA AG, Hamburg, Germany, and Department of Psychology, City University, London, UK; [10]Department of Clinical Psychology, VU University Amsterdam, and Department of Clinical Psychological Science, Maastricht University; [11]Department of Epidemiology and Biostatistics and EMGO Institute for Health and Care Research, VU University Medical Center; [12]Centre for Psychological Services Research, University of Sheffield, Sheffield, UK; [13]University of Wollongong, Wollongong NSW, Australia; [14]School of Nursing, Midwifery and Social Work, University of Manchester; [15]College of Medicine, Fu-Jen Catholic University, Taiwan

Abstract

Objective To assess how initial severity of depression affects the benefit derived from low intensity interventions for depression.

Design Meta-analysis of individual patient data from 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 by means of written materials and limited professional support, and internet delivered interventions).

Main outcome measures Depression outcomes (measured with the Beck Depression Inventory or Center for Epidemiologic Studies Depression Scale), and the effect of initial depression severity on the effects of low intensity interventions.

Results Although patients were referred for low intensity interventions, many had moderate to severe depression at baseline. We found a significant interaction between baseline severity and treatment effect (coefficient –0.1 (95% CI –0.19 to –0.002)), 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 (equivalent to an additional drop of around one point on the Beck Depression Inventory for a one standard

Correspondence to: P Bower peter.bower@manchester.ac.uk

deviation increase in initial severity) was small and may not be clinically significant.

Conclusions The data suggest that patients with more severe depression at baseline show at least as much clinical benefit from low intensity interventions as less severely depressed patients and could usefully be offered these interventions as part of a stepped care model.

Introduction

Depression is a major cause of disability among populations worldwide,[1] and effective management is a key challenge for healthcare systems. In response, some have recommended a stepped care approach,[2] and this has been adopted as the basis for depression services in the UK.[3] In stepped care, a large proportion of patients are first treated with "low intensity" psychological interventions,[4] which are generally based on cognitive behavioural therapy and delivered via written materials or information technology with limited professional guidance (see box 1). Evidence suggests low intensity interventions provide significant clinical benefit.[5][6] In stepped care, conventional high intensity interventions (such as 12–16 sessions of therapist led cognitive behavioural therapy) are offered only to those who fail to respond to initial low intensity interventions, or to 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 intensity and high intensity psychological therapy is initial severity of depression. However, the 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 and patient experience in depression services.[7]

Variables which predict response to interventions are described as moderators of treatment effect.[8] Despite the existence of a relatively large literature on the effectiveness of low intensity interventions,[5][9-12] there is relatively little rigorous evidence on the critical clinical question of whether initial severity moderates effectiveness of low intensity interventions—that is, do more severely ill patients show better or worse treatment effects? Study level meta-analyses[12][13] of these relationships lack precision and are vulnerable to ecological bias.[14] Individual studies often report moderators as secondary analyses, but their yield has been limited by scarcity, selective reporting,[15] inappropriate methods,[8][16] and low power, as sample sizes required to achieve power to detect moderators are potentially very high.[17] 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.[18]

We describe an individual patient data meta-analysis of depression severity as a moderator of the effect of low intensity interventions in depression,[19] 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.[5][6][9-12][20-23] We updated these with additional searches of the Cochrane Library in July 2011 (see "Additional resources" file on bmj.com 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 commonly 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 of contact). We excluded self help groups and any low intensity intervention 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.[18] The study protocol is available in the "Additional resources" file on bmj.com..

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 datasets were consistent with published data.

Measure standardisation

Almost all studies either used the Beck Depression Inventory (BDI)[24] or Center for Epidemiologic Studies Depression Scale (CES-D)[25] as the main depression outcome. We report scores on these scales for descriptive purposes, converting one trial using the Clinical Outcomes in Routine Evaluation Outcome Measure (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

> **Box 1: Stepped care**
>
> Stepped care is a system of delivering health technologies, where the most effective yet least resource intensive treatment is delivered to patients first
>
> In depression care, conventional psychological therapies such as cognitive behavioural therapy (so called high intensity therapies involving 12–16 sessions from an experienced practitioner) are effective, but demand outstrips supply, leading to long waiting lists
>
> Less resource intensive versions, delivered via books and information technology with limited support and guidance from a professional, have been developed (so called low intensity interventions)
>
> Stepped care is intended to enhance efficiency by providing low intensity interventions to a proportion of depressed patients in the first instance, before providing higher intensity interventions to those who do not improve with the low intensity interventions. Stepped care is best seen as the product of two simple principles:
>
> 1. The principle of "least burden"—Effective low intensity interventions are offered to patients first, and high intensity interventions offered only to patients who are at risk to self or others, have a history of treatment failure, or do not improve from initial treatment
>
> 2. The principle of "scheduled review"—This is required so that patients can "step up" to more intensive interventions, or change to another intervention within the same step, if they fail to meet consensus criteria for improvement or recovery. Scheduled reviews use objective outcome measures to assist decision making

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 we 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. 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 1000 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 imputation algorithms was verified with time series and autocorrelation plots of the worst linear function.[30 31] We tested whether baseline variables (study, group allocation, age, sex, and baseline depression) predicted missing data to test the assumptions underlying imputation. We also conducted a sensitivity analysis using only cases with available data.

Analysis

As individual patient data meta-analyses are vulnerable to publication bias from a number of sources,[32] 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 data were unavailable. We present descriptive statistics on study characteristics (including quality, in terms of concealment of allocation, reporting of intention to treat analysis, and attrition rates of <20%). We also assessed the potential for publication bias using funnel plots, in line with published recommendations.[32] 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").[14 18] In certain situations these last two analyses give identical results, although they differ under conditions such as "covariate heterogeneity" (that is, the variation in the covariate within each study).[14]

In this study we used the one step analysis, which is the most logistically demanding but which 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.[33 34] Appropriate mixed effects models (with fixed trial-specific intercepts for the interaction, a random treatment effect, and fixed trial-specific effects for baseline) were used to synthesise the patient level data and estimate the variances between and within studies, fitting the interaction as a continuous variable.[35] We also repeated these analyses with different meta-analytic models (random trial intercept; random treatment effect; fixed trial-specific effects for baseline). We used Stata v12.1 and a restricted maximum likelihood algorithm with the "xtmixed" command.[36 37] Heterogeneity was assessed using the I^2 statistic.[38] For cluster randomised studies, we adjusted appropriately.[39] 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 (that is, 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 two 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 study quality. Although the comprehensive Cochrane risk of bias tool[40] is widely used, we needed a measure of quality that could be used in the quantitative analysis. We chose a dichotomous measure based on allocation concealment, as this is the aspect of quality most consistently associated with treatment effect,[41 42] is particularly relevant when outcomes are subjective,[43] and because other measures included in the risk of bias tool, such as blinding, are generally less useful in trials of psychological therapy because the conditions for blinding are so rarely met and most outcomes are self reported. Allocation concealment was judged as adequate or inadequate according to the relevant section from the Cochrane risk of bias tool.

We also coded the types of low intensity interventions: internet versus written forms, and "guided" (low intensity interventions with limited support by a health professional) versus "unguided" forms (used by the patient alone). An additional post hoc secondary analysis explored whether the main result was influenced by the outcome measure used (BDI or CES-D).

Results

Figure 1⇓ shows the process of study selection for our review. We excluded six potentially eligible studies because numbers of participants were below 50, five because they were published before 2000, and four 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 (SE 0.8), P=0.007, fig 2⇓). We gained access to data from 16 (55%) of these comparisons, with data unavailable either because of no response from authors (n=8), clashes with their own planned analyses (n=4), or ethical issues with sharing data (n=1). A small number of individual cases were dropped because of missing baseline age or depression scores, leaving 2470 unique cases, with 77% reporting data at first follow-up. Group allocation had the strongest association with missing follow-up data, with patients in the usual care group less likely to have missing outcome data. Such patterns of missing data might be expected to result in an inflation of the overall effect (if missing data was associated with poor outcomes), but the effect on the interaction is difficult to predict.

Available and unavailable data

Data on study characteristics and design are detailed in the "Additional resources" file on bmj.com. We compared available and unavailable studies on population, intervention, quality, and outcome data (see table⇓). Studies were similar in recruitment procedures, although available studies were less likely to involve patients with a diagnosis of depression or health technologies delivered via information technology, but were more likely to involve support from a health professional. Available studies met more quality criteria, had a slightly 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 depression of patients included in these trials. Six studies (38%) had a maximum ceiling for inclusion. Assessment of mean depression scores at baseline showed that many patients had appreciable symptoms (see fig 3⇓). For the BDI score (range 0–63), a score of 10–16 indicates mild depression, 17–29 indicates moderate depression, and ≥30 indicates severe depression: the studies' mean scores were 19–21,[44] 21,[45] 22,[46] 23–24,[47] 23–28,[48] 26,[49] 27,[50] 27–28,[51] and 29.[52] For the CES-D score (range 0–60), a score of ≥16 indicates a probable depressive illness, and the studies' mean scores ranged from 13 in a trial focussed on subthreshold symptoms[53] to 21–22,[54] 30,[55] and 32.[56]

In terms of other characteristics of the patients, comparisons are limited by the data presented and reflect study inclusion criteria, but generally two thirds to three quarters of patients were women, with mean ages 35–45 years, and with rates of university education ranging from 20% to 65%. In terms of treatment history, rates of current antidepressant use (where

reported) ranged from 19% to 69%, and between 38% and 67% reported a previous treatment for depression.

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

The overall standardised estimate of the main effect of low intensity interventions was −0.42 (95% confidence interval −0.55 to −0.29, I^2=2.9% (0.5% to 15%)). This would be equivalent to an additional drop of around four or five points on both BDI and CES-D scores, over and above the change in the controls. There was no evidence that this main effect varied by age, sex, intervention type, or study quality. When a term was added to assess the interaction, we found a significant negative interaction between baseline severity and treatment effect (interaction coefficient −0.1 (−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 one standard deviation, which would be equivalent to an additional drop of around one point on both BDI and CES-D for a one standard deviation increase in initial severity, an effect which may not be clinically significant. The interpretation of the main result is outlined in clinical terms in box 2.

Figure 4⇓ shows the estimates of the interactions at the level of the individual studies. The estimate was similar when conducted on available data without imputation (−0.12 (95% confidence interval −0.22 to −0.02)) and was not sensitive to variation in the meta-analytic model specified or the different measures included in the trials (BDI or CES-D score).

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 patients with more severe depression at presentation. When data were analysed in terms of five 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 coefficient −0.07 (95% confidence interval −0.34 to 0.21)).

Are the results sensitive to types of low intensity interventions?

The moderating effect of initial depression was larger in patients in the studies that used internet based low intensity interventions, compared with the studies that used written interventions, but the difference was not statistically significant (interaction coefficient −0.09 (−0.31 to 0.12)). The moderating effect of initial depression was also greater in patients who used unguided low intensity interventions, compared with those who used

Box 2: Clinical scenario

Patients attending primary care and considered eligible for psychological therapy for depression may present with a Beck Depression Inventory (BDI) score of around 25 (out of a maximum of 63), indicating moderate severity of depression. After three to six months in usual primary care, without any intervention, such patients might be expected to reduce their score on average by four points to around 21, still indicative of moderate depression.

If such patients were referred to a low intensity intervention, they might be expected to display an additional reduction of four points on average, over and above this natural change over time, to a score of around 17, indicative of milder depression.

The evidence presented in this paper would suggest that patients who present with more severe problems (such as a presenting score of 35) would show an additional drop of around one point (a total reduction of around five points) compared with those with an initial score of 25.

The results are displayed visually below. The horizontal axis shows initial severity of depression, and the vertical axis shows severity at follow-up. As can be seen from fig 5⇓, patients in the low intensity intervention group consistently demonstrate lower severity of depression at follow-up than usual care patients. These lower scores are evident across the entire range of initial depression severity (that is, the lines never cross). The additional benefit shown by patients treated with low intensity interventions increases as initial severity increases (that is, the vertical distance between the lines increases as initial depression severity increases). However, the magnitude of this divergence is relatively small and is unlikely to be clinically significant.

The data illustrate that:

(*a*) low intensity interventions provide clinically significant benefits over usual care

(*b*) patients with more severe problems show greater levels of benefit from low intensity interventions, although the magnitude of that additional benefit is modest and may not be clinically significant.

Although patients with more severe depression show greater benefits over usual care, their initial high scores mean that they are more likely to continue to show clinically important levels of distress after low intensity interventions and may require additional care.

guided interventions, but again the difference was not significant (interaction coefficient −0.07 (−0.30 to 0.15)).

Discussion

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 much clinical benefit from the interventions as less severely ill patients. We did not find evidence that the main result was dependent on characteristics of the studies, or the interventions, or major analytical assumptions.

Strengths and weaknesses of the study

Although generally considered as a gold standard, meta-analyses using individual patient data are potentially vulnerable to publication bias (selective publication of significant results in primary studies), reviewer selection bias (selective identification of relevant datasets of individual patient data) and availability bias (selective access to individual patient datasets once identified). The funnel plot suggested some 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%.[32] We were able to access just over half of the eligible studies and patients. 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. It should be noted that the studies available to the review met more of our quality criteria (allocation concealment, intention to treat analyses, and low attrition) than studies where data were unavailable (see table⇓), with over 80% reporting adequate concealment of allocation.

As noted previously, it is possible that patients with 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 that used the BDI score included 430 patients (nearly a third of the total) with scores >30 (indicating severe depression), which shows 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 of up to 40 on the two outcome measures. The analysis assumes equivalence in the clinical meaning of change at different levels of initial severity, such that the impact of a reduction in scores for a patient who initially scores 30 is the same as that for a patient scoring 16. This assumption is conventional in trial analyses.

Although our results were robust to a range of sensitivity analyses, it should be noted that the tests of three way interactions (such as tests of whether the interaction of initial severity and outcome differed in studies of different quality) lacked precision.

Strengths and weaknesses in relation to other studies

There are no comparable analyses in the literature of low intensity interventions for depression. Thirteen comparisons in the total dataset included some form of secondary analysis of moderators (see table of study characteristics in "Additional resources" on bmj.com), although the variables tested and the analytical techniques used varied widely, and not all explored severity. Of those examining initial severity of depression, four comparisons suggested similar results in less and more severely ill patients,[54 57 58] one reported a greater benefit in less severely ill patients,[52] and the rest reported that more severely ill patients showed greater benefits.[51 55 59] 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 analysis 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.[13] 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 lack of clinically meaningful differences in treatment effects related to baseline severity would suggest that it is legitimate to include low intensity interventions in the first step of a stepped care system and to encourage most patients to use them as the initial treatment option, even when initial severity of depression is high. Clearly some patients will not find such interventions useful, and it would seem sensible to continue to refer severe cases to more intense psychological intervention or pharmacological management until further evidence is generated confirming our findings. The current data 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 modelled the impact of initial severity only 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 (see box 1) 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.[60] It is possible that immediate provision of high intensity interventions to patients with more severe depression would be more cost effective than initial use of low intensity interventions followed by high intensity therapy. Secondly, it is possible that initial experience with low intensity interventions (especially if unsuccessful) could act as a barrier to further treatment. Data to explore either of these hypotheses are not available at present, and this remains an important research question for the future.

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 patients' needs and preferences are important, but the evidence demonstrating a relationship between preferences and outcome is inconsistent.[61 62] The effects of preferences could in principle be tested in a similar way to the current analysis if baseline data were reported consistently.[62]

Unanswered questions and future research

Our results show that some of the concerns about examination of moderators in clinical trials (especially those around sample size) can be overcome through collaborative meta-analysis of individual patient data. 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.

Our 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[62] 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,[63] 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.

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 Davison G. Stepped care: doing more with less? *J Consult Clin Psychol* 2000;68:580-5.
3 National Institute for Health and Clinical Excellence. *Depression: the treatment and management of depression in adults (update)*. NICE, 2010. www.nice.org.uk/nicemedia/pdf/Depression_Update_FULL_GUIDELINE.pdf.
4 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 University Press, 2010.
5 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:e13196.
6 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;40:1943-57.
7 Richards D, Bower P, Pagel C, Weaver A, Utley M, Cape J, et al. Delivering stepped care for depression: an analysis of implementation in routine practice. *Implementation Science* 2012;7:3.
8 Kraemer H, Wilson G, Fairburn C, Agras W. Mediators and moderators of treatment effects in randomized clinical trials. *Arch Gen Psychiatry* 2002;59:877-83.
9 Andersson G, Cuijpers P. Internet-based and other computerized psychological treatments for adult depression: a meta-analysis. *Cognitive Behaviour Therapy* 2010;38:196-205.
10 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.
11 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;10:1-168.
12 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;37:1217-28.
13 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;78:668-80.
14 Simmonds M, Higgins J. Covariate heterogeneity in meta-analysis: criteria for deciding between meta-regression and individual patient data. *Stat Med* 2011;26:2982-99.
15 Sun X, Briel M, Busse J, You J, Akl E, Mejza F, et al. The influence of study characteristics on reporting of subgroup analyses in randomised controlled trials: systematic review. *BMJ* 2012;342:d1569.
16 Sun X, Briel M, Busse J, You J, Akl E, Mejza F, et al. Credibility of claims of subgroup effects in randomised controlled trials: systematic review. *BMJ* 2012;344:e1553.
17 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;57:229-36.
18 Riley R, Lambert P, Abo-Zaid G. Meta-analysis of individual participant data: rationale, conduct, and reporting. *BMJ* 2010;340:c221.

What is already known on this topic

To better manage depression in the community, many services seek to provide simple forms of psychological therapy (so called low intensity interventions) to depressed patients

Patients with more severe depression may be less suitable for low intensity interventions, but evidence is lacking about which patients should receive low intensity interventions

What this study adds

This meta-analysis of individual patient data from 16 datasets found no clinically meaningful differences in treatment effects between more and less severely ill patients receiving low intensity interventions

Patients with more severe depression can be offered low intensity interventions as part of a stepped care model

19 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;14:2057-79.

20 Anderson L, Lewis G, Araya R, Elgie R, Harrison G, Proudfoot J, et al. Self-help books for depression: how can practitioners and patients make the right choice? *Br J Gen Pract* 2005;55:387-92.

21 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;51:838-45.

22 Cuijpers P. Bibliotherapy in unipolar depression: a meta-analysis. *J Behav Ther Exp Psychiatry* 1997;28:139-47.

23 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;6:1-89.

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, Connell J, Barkham M, Margison F, McGrath G, Mellor-Clark J, et al. Towards a standardised brief outcome measure: psychometric properties and utility of the CORE-OM. *Br J Psychiatry* 2002;180:51-60.

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;45:153-66.

28 Gadbury G, Coffey C, Allison D. Modern statistical methods for handling missing repeated measurements in obesity trial data: Beyond LOCF. *Obesity Rev* 2003;4:175-84.

29 Donders A, van der Heiden G, Stijnen T, Moons K. Review: a gentle introduction to imputation of missing values. *J Clin Epidemiol* 2012;59:1087-91.

30 Schafer J. *Analysis of incomplete multivariate data* . Chapman and Hall, 1997.

31 Stata Press. *Multiple-imputation reference manual* . Stata Press, 2012.

32 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.

33 Stewart L, Parmar M. Meta-analysis of the literature or of individual patient data: is there a difference? *Lancet* 1993;341:418-22.

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

35 Whitehead A. *Meta-analysis of controlled clinical trials* . John Wiley & Sons, 2002.

36 StataCorp. Stata statistical software. [11.1]. StataCorp, 2009.

37 Kontopantelis E, Reeves D. A short guide and a forest plot command (ipdforest) for one-stage meta-analysis. *Stata Journal* (forthcoming).

38 Higgins J, Thompson S. Quantifying heterogeneity in a meta-analysis. *Stat Med* 2002;21:1539-58.

39 Sutton A, Kendrick D, Coupland C. Meta-analysis of individual- and aggregate-level data. *Stat Med* 2008;27:651-69.

40 Higgins J, Altman D, Gøtzsche P, Jüni P, Moher D, Oxman A, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised studies. *BMJ* 2011;343:d5928.

41 Schulz K, Chalmers I, Hayes R, Altman D. Empirical evidence of bias: dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995;273:408-12.

42 Pildal J, Hróbjartsson A, Jørgensen K, Altman D, Gøtzsche P. Impact of allocation concealment on conclusions drawn from meta-analyses of randomized trials. *Int J Epidemiol* 2007;36:847-57.

43 Wood L, Egger M, Gluud L, Schulz K, Jüni P, Altman D, et al. Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. *BMJ* 2012;336:601-5.

44 Bilich L, Deane F, Phipps A, Barisic M, Gould G. Effectiveness of bibliotherapy self-help for depression with varying levels of telephone helpline support. *Clin Psychol Psychother* 2008;15:61-74.

45 Andersson G, Bergstrom J, Hollandare F, Carlbring P, Kaldo V, Ekselius L. Internet-based self-help for depression: randomised controlled trial. *Br J Psychiatry* 2005;187:456-61.

46 Vernmark K, Lenndin J, Bjärehed J, Carlsson M, Karlsson J, Öberg 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;48:368-76.

47 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;53:764-70.

48 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;33:449-61.

49 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;35:1633-43.

50 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. *J Med Internet Res* 2009;11:e15.

51 De Graaf L, Gerhards S, Amtz A, Riper H, Metsemakers J, Evers M, et al. Clinical effectiveness of online computerised cognitive behavioural therapy without support for depression in primary care: randomised trial. *Br J Psychiatry* 2009;195:73-80.

52 Lovell K, Bower P, Richards D, Barkham M, Sibbald B, Roberts C, 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;8:91.

53 Willemse G, Smit F, Cuijpers P, Tiemens B. Minimal contact psychotherapy for sub-threshold depression in primary care: randomised trial. *Br J Psychiatry* 2004;185:416-21.

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

55 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.

56 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.

57 Proudfoot J, Ryden C, Everitt B, Shapiro D, Goldberg D, Mann A, 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.

58 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;36:325-33.

59 Levin W, Campbell D, McGovern K, Gau J, Kosty D, Seeley J, et al. A computer-assisted depression intervention in primary care. *Psychol Med* 2010;41:1373-83.

60 Bower P, Gilbody S. Stepped care in psychological therapies: access, effectiveness and efficiency: narrative literature review. *Br J Psychiatry* 2005;186:11-7.

61 King M, Nazareth I, Lampe F, Bower P, Chandler M, Morou M, et al. Impact of participant and physician intervention preferences on randomised trials: a systematic review. *JAMA* 2005;293:1089-99.

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

63 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.

Table

Table 1| Comparison of available and unavailable studies. Values are numbers (percentages) unless stated otherwise

| Factor	Unavailable (n=13)	Available (n=16)
Recruitment via screening (versus referral)	10 (77)	13 (81)
Depression diagnosis confirmed	6 (46)	2 (13)
Computerised delivery (versus bibliotherapy)	12 (92)	10 (63)
Guided minimal intervention	6 (46)	12 (75)
Mean quality (0–3)*	1.9	2.3
Mean baseline number	145	156
Pooled effect size (95% CI)	−0.47 (−0.27 to −0.68)	−0.39 (−0.26 to −0.52)

*Number of quality criteria on which studies were judged as adequate (criteria were adequate concealment of allocation, reporting of intention to treat analysis, and <20% attrition).
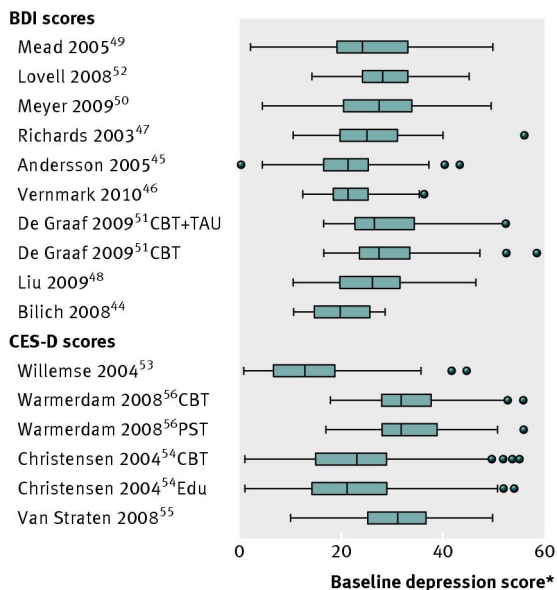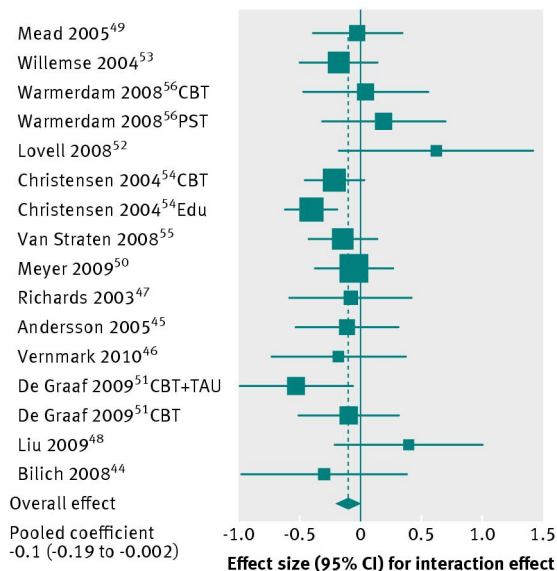
Figures
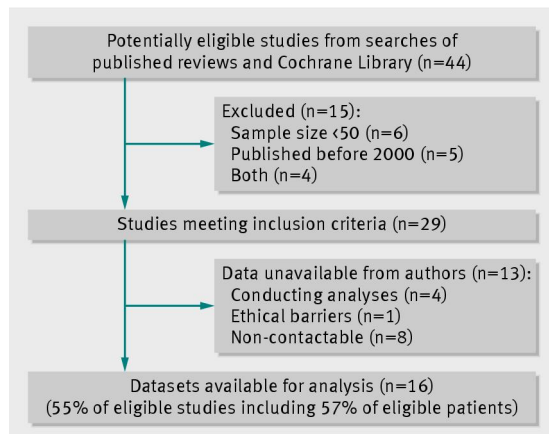
Fig 1 Inclusion of studies in the review
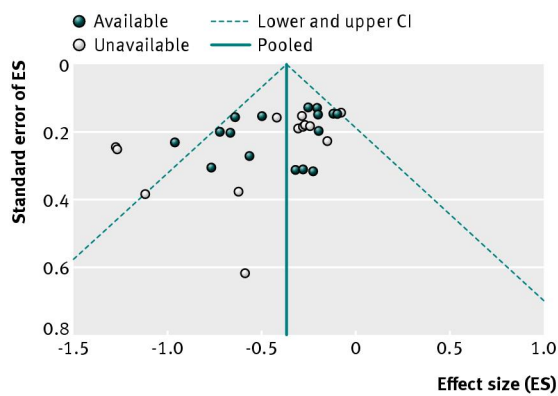
Fig 2 Funnel plot of studies included in analysis with pseudo 95% confidence intervals. Egger's regression intercept –2.39 (SE 0.8), P=0.007

BDI scores
Mead 2005[49]
Lovell 2008[52]
Meyer 2009[50]
Richards 2003[47]
Andersson 2005[45]
Vernmark 2010[46]
De Graaf 2009[51]CBT+TAU
De Graaf 2009[51]CBT
Liu 2009[48]
Bilich 2008[44]
CES-D scores
Willemse 2004[53]
Warmerdam 2008[56]CBT
Warmerdam 2008[56]PST
Christensen 2004[54]CBT
Christensen 2004[54]Edu
Van Straten 2008[55]

Baseline depression score*

*BDI score range 0-63
CES-D score range 0-60
CBT=cognitive behavioural therapy
TAU=treatment as usual
PST=problem solving therapy
Edu=education

Fig 3 Baseline severity data of studies included in the review. Box and whisker plots show median, interquartile range, minimum and maximum scores, and outliers

Study

Mead 2005[49]
Willemse 2004[53]
Warmerdam 2008[56]CBT
Warmerdam 2008[56]PST
Lovell 2008[52]
Christensen 2004[54]CBT
Christensen 2004[54]Edu
Van Straten 2008[55]
Meyer 2009[50]
Richards 2003[47]
Andersson 2005[45]
Vernmark 2010[46]
De Graaf 2009[51]CBT+TAU
De Graaf 2009[51]CBT
Liu 2009[48]
Bilich 2008[44]
Overall effect
Pooled coefficient
-0.1 (-0.19 to -0.002)

Effect size (95% CI) for interaction effect

CBT=cognitive behavioural therapy
TAU=treatment as usual
PST=problem solving therapy
Edu=education

Fig 4 Forest plot of interactions between baseline severity of depression and effect of low intensity interventions
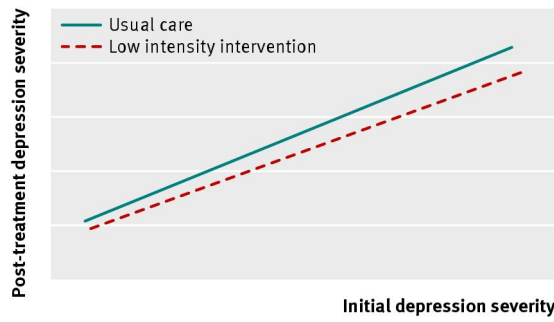
Fig 5 How baseline severity of depression moderates the effect of low intensity interventions on depression