ORIGINAL ARTICLE

Modelling the cost-effectiveness of pharmacotherapy compared with cognitive–behavioural therapy and combination therapy for the treatment of moderate to severe depression in the UK

L. Koeser[1]*, V. Donisi[2], D. P. Goldberg[1] and P. McCrone[1]

[1] *Institute of Psychiatry, King's College London, London, UK*
[2] *Department of Public Health and Community Medicine, Section of Psychiatry, University of Verona, Verona, Italy*

Background. The National Institute of Health and Care Excellence (NICE) in England and Wales recommends the combination of pharmacotherapy and psychotherapy for the treatment of moderate to severe depression. However, the cost-effectiveness analysis on which these recommendations are based has not included psychotherapy as monotherapy as a potential option. For this reason, we aimed to update, augment and refine the existing economic evaluation.

Method. We constructed a decision analytic model with a 27-month time horizon. We compared pharmacotherapy with cognitive–behavioural therapy (CBT) and combination treatment for moderate to severe depression in secondary care from a healthcare service perspective. We reviewed the literature to identify relevant evidence and, where possible, synthesized evidence from clinical trials in a meta-analysis to inform model parameters.

Results. The model suggested that CBT as monotherapy was most likely to be the most cost-effective treatment option above a threshold of £22 000 per quality-adjusted life year (QALY). It dominated combination treatment and had an incremental cost-effectiveness ratio of £20 039 per QALY compared with pharmacotherapy. There was significant decision uncertainty in the probabilistic and deterministic sensitivity analyses.

Conclusions. Contrary to previous NICE guidance, the results indicated that even for those patients for whom pharmacotherapy is acceptable, CBT as monotherapy may be a cost-effective treatment option. However, this conclusion was based on a limited evidence base, particularly for combination treatment. In addition, this evidence cannot easily be transferred to a primary care setting.

Introduction

Depression is the most common psychiatric disorder. In 2007, 1.24 million people were estimated to suffer from this condition in England, which resulted in health and social care costs of £1.7 billion. By 2026, the number of people and costs associated with depression are projected to rise to 1.45 million and £3 billion, respectively (McCrone *et al.* 2008). Treatment for depression in the England and Wales is guided by a stepped-care model (National Institute of Health and Care Excellence, 2009). Less intrusive and costly treatments such as guided self-help, physical exercise or computerized cognitive–behavioural therapy (CBT) are recommended for patients with subthreshold or mild symptoms. Access to more intensive treatments (i.e. pharmacotherapy, psychotherapy and combination therapy) is indicated for patients who do not respond to or decline these options, or those presenting with more severe types of depression. In order to make best use of limited health care budgets, the evaluation of the relative cost-effectiveness of these three treatment alternatives has been a chief concern in the UK. Many studies assessing different types of antidepressants exist but comparisons between difficult treatment classes are less common. The fundamental question in this context is whether the added costs of psychotherapy, singly or in combination with pharmacotherapy, due to the more extensive contact with clinicians, are outweighed by the potential benefits over pharmacotherapy.

* Address for correspondence: L. Koeser, Health Service and Population Research Department, Institute of Psychiatry, King's College London, Box P024, De Crespigny Park, London SE5 8AF, UK.
(Email: leonardo.koeser@kcl.ac.uk)

Current guidance by the National Institute for Health and Care Excellence (NICE) recommends the provision of a combination of pharmacotherapy and CBT or interpersonal psychotherapy for both moderate and severe depression (National Institute of Health and Care Excellence, 2009). However, this recommendation is based on an economic analysis that does not include psychotherapy in the form of a monotherapy as a treatment alternative. The National Institute of Health and Care Excellence (2009) guidance development group (GDG) justified this exclusion by noting that the 'clinical evidence review showed no overall superiority for CBT alone on treatment outcomes' and that it had significantly higher treatment costs. Yet, elsewhere the GDG states that 'it was not possible to identify a benefit of adding antidepressants to CBT', and that 'CBT alone was found to be better than antidepressant alone when compared with combined treatment'. In addition, in practice many patients in the UK do not appear to receive recommended treatments because of supply constraints in the provision of psychotherapy (Gyani *et al.* 2012). For these reasons, this study aimed to update and refine the available evidence on the comparative cost-effectiveness of treatments for depression to inform decision making in a secondary care setting. In particular, besides comparing pharmacotherapy and combination treatment we also included CBT monotherapy in the economic evaluation.

Method

Patient population and comparators

The target population of our decision analytic model was adults with moderate or severe major depressive disorder (MDD) according to cut-off scores on two common depression rating scales: the 17-item Hamilton Depression Scale (HAMD-17) (scores $\geqslant 14$) and the Beck Depression Inventory (scores $\geqslant 17$) (American Psychiatric Association, 2000). Given the amount of available data, we chose to combine the available evidence base using a decision tree model. We included studies in which any antidepressant medication, face-to-face CBT or cognitive therapy (CT) and/or combination treatment were compared. We excluded other types of psychotherapies as they are not commonly available in the UK National Health Service (NHS) and because to date less empirical evidence on them is available. We did not consider other forms of delivering psychotherapy such as group therapy or computerized CBT. We modelled treatment in a secondary care setting because the vast majority of patients in eligible studies recruited patients in this context. We only considered first-line treatment and did not allow for treatment augmentation or switching during the acute treatment phase.

Model structure

None of the clinical trials relevant for our model followed up patients for more than 24 months after the end of the acute treatment phase. Therefore, we compared the cost-effectiveness of antidepressants, CBT and combination therapy over a 27-month time horizon consisting of a 3-month acute treatment phase, and follow-up at 12 months and 24 months after the end of acute treatment. We distinguished between three post-treatment clinical states or events: remission (or full response); response (or partial remission); and non-response.

Premature termination of treatment is common in depression. In the model, patients were assumed to discontinue treatment because of positive reasons (i.e. remission or the perception thereof) or negative ones (i.e. no improvements in symptoms and/or side effects). Patients who remitted or responded to treatment were thought to be at risk of relapse in the follow-up phases of the model. Fig. 1 shows the possible transition pathways of the model.

Event probabilities

Where possible, we obtained data for model parameters from randomized controlled trials (RCTs) because they are believed to minimize the risk of bias (National Institute of Health and Care Excellence, 2009). A regularly updated database by Cuijpers *et al.* (2008) contains a catalogue of RCTs that compare the effects of psychological treatments (both singly and in combination) in adults with depressive disorders with a control intervention. V.D. and L.K. independently screened the January 2013 version of this database to identify relevant head-to-head comparisons. We included English-language studies only.

We extracted data on remission, response and dropout rates if the study reported these outcomes at completion of the acute treatment phase. We considered patients with a score of 7 or less on the HAMD-17 to be in remission, those reaching a score between 8 and 13 to be responders and those with a score of 14 or above as non-responders (National Institute of Health and Care Excellence, 2009). However, we allowed for a ±1 point margin in the cut-off definitions due to slight variations between studies. We extracted the data on an intention-to-treat basis. For consistency with National Institute of Health and Care Excellence (2009) guidelines, we only extracted data on relapse rates from trials that incorporated some form of maintenance treatment, i.e. pharmacotherapy had to be continued beyond the end of the acute phase and 'booster' sessions had to available to patients both in the monotherapy and combination therapy arms. In light of the small number of studies that met this criterion, we allowed for any definition of relapse and allowed for studies
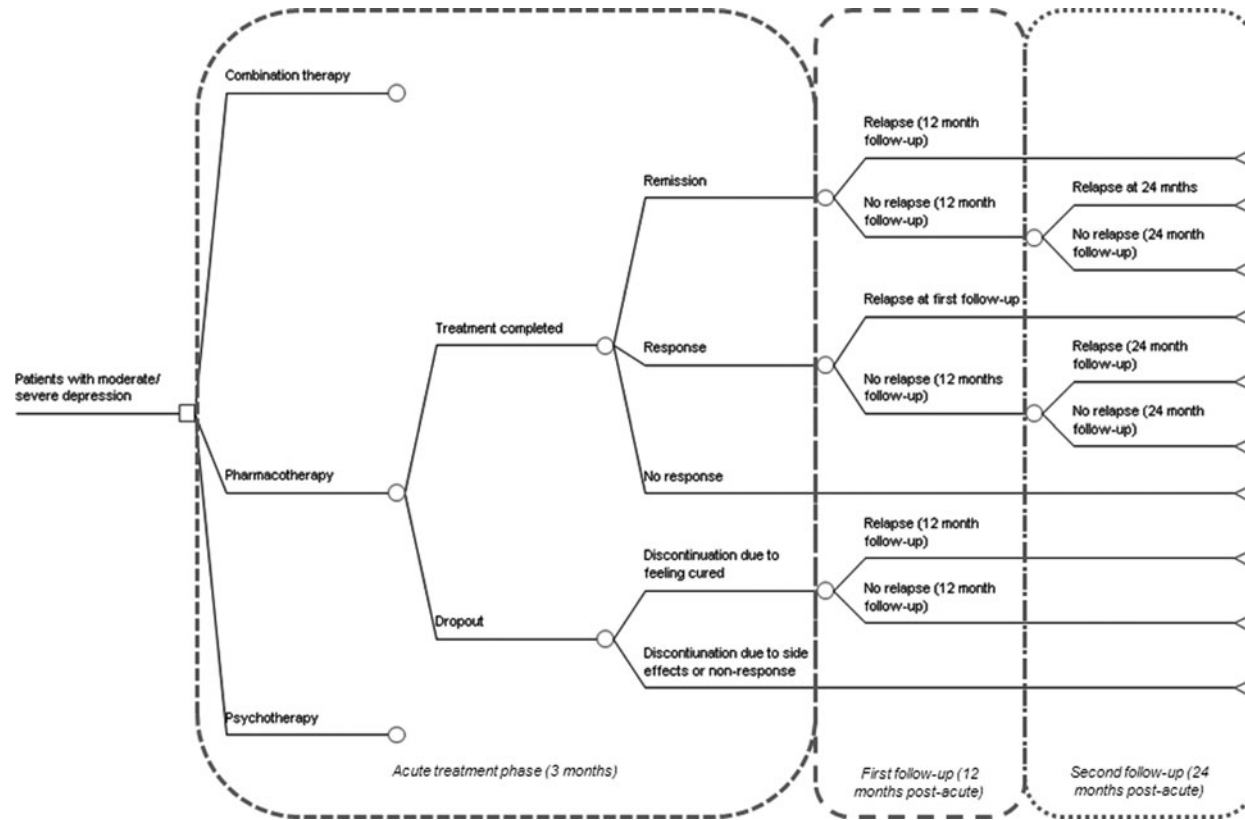
Fig. 1. Partial decision tree structure (treatment pathways in combination therapy and psychotherapy are the same as those for pharmacotherapy).

3022   *L. Koeser et al.*

that included some patients that may have been responders but not remitters according to our definition. In practice, in all included trials pharmacotherapy was discontinued at 6 or 12 month after remission. Based on our knowledge of the disease area and for sake of clarity, we considered it to be most appropriate to take a 'worst-case scenario' approach to missing data (Higgins & Green, 2008). In other words, we assumed that patients with a missing endpoint assessment in the acute and follow-up phases would be in the least favourable health status (i.e. non-response or relapse) and if it was unclear at what point the patient dropped out of the study during the post-acute follow-up, we assumed that this occurred before the first-follow up. There appeared to be some ambiguities in how the data were reported in the follow-up studies. Given the importance of relapse rates for the results of the model, we individually comment on our approach to relapse rate data extraction in online Supplementary Appendix S3.

We were unable to identify randomized trials that compared relapse rates among patients responding to treatment in isolation from remitters according to our definition of these subgroups. For this reason, over the first 12 months of follow-up, for the pharmacotherapy arm in our model we used the relapse rate of patients with a HAMD-17 score between 8 and 13 in the trial by Kuyken *et al.* (2008) who were treated with antidepressants. Conditional on not having relapsed until this point, we assumed that the relapse rate among pharmacotherapy treatment responders over the second 12-month follow-up was the same as among remitters. It appeared plausible that the direction of relative differences in relapse rates between treatments in terms of protecting against relapse would be the same among patients who were in remission after the acute treatment phase as for those responding to treatment. In absence of any direct evidence to support this belief or data on the relative magnitude of effects, we adopted a conservative approach aimed to reflect this notion in our model without unduly favouring any of the three interventions. Specifically, we assumed that the risk differences in relapse rates between the three interventions would be the same as among remitters but multiplied by a discount factor with mean 0.5 and S.D. 0.3, i.e. reduced by 50% on average.

A priori, there were clinical reasons to believe that some heterogeneity in observed treatment effects would be present. Besides the fact that different types of antidepressants were included in our comparisons, there is considerable variability in how both psychotherapy and pharmacotherapy are implemented in trials. Therefore, we synthesized the evidence using random-effects meta-analysis which does not assume the presence of a common effect across all studies

(Borenstein *et al.* 2009; Riley *et al.* 2011). We adapted a Bayesian network meta-analysis framework proposed by Dias *et al.* (2013*a*) that accounts for both direct and indirect treatment effects and correlations between arms within a trial. To better reflect the uncertainty resulting from heterogeneity between trials in a decision-making context, we used predictive rather than posterior distributions for baseline rates and treatment effects in our model (Dias *et al.* 2013*b, c*). In other words, we modelled the uncertainty surrounding a hypothetical future 'roll out' of the interventions given the between-study variance rather than the uncertainty surrounding the average presumed underlying treatment effects. We applied vague prior distribution for baseline rates and weakly informative t family prior to treatment effects (Gelman *et al.* 2008). For between-study variances, on the other hand, we used informative prior distributions based on a review by Turner *et al.* (2012) to stabilize our estimates. Put differently, when there were few studies available to inform the estimate of between-study variance, such as in the follow-up phase, we assumed that variance between studies in our meta-analysis would be relatively similar to that in comparable meta-analyses in the literature rather than relying on the limited amount of existing data, whereas we allowed the data to 'dominate' when sufficient evidence was available. We used WinBUGS to run our analyses. The code which allows the replication of the entire decision model including these meta-analyses can be found in online Supplementary Appendix S4.

The literature contained little systematic evidence on the disease course of patients who discontinued depression treatment. Data by Radhakrishnan *et al.* (2013) as well as expert opinion elicited by the National Institute of Health and Care Excellence (2009) and Sado *et al.* (2009) suggested that approximately 20% of patients discontinue treatment due to recovery and this figure was used. We assumed that, regardless of initial treatment assignment, relapse rates at 12-month follow-up among patients who discontinued treatment because of feeling cured were equal to those of patients treated with a placebo during the acute phase in a study by Jarrett *et al.* (2000). In addition, we assumed that patients in this subgroup who did not relapse during the first 12-month follow-up were not at risk of relapse over the remaining time horizon of the model.

Health-related quality of life

We quantified the health benefits of the interventions using quality-adjusted life years (QALYs). This approach assigns a preference-based weight, usually between 1 (representing full health) and 0 (representing death) to health states in an attempt to quantify the

relative value of quality of life therein. This value is multiplied by the length of time spent in that health state to yield QALYs (Malek, 2000). The EuroQol 5 Dimensions (EQ-5D) is the instrument currently preferred by NICE to derive the preference weights for health states in adults (National Institute of Health and Care Excellence, 2013). However, a review of the literature suggested that no published evidence was available mapping EQ-5D scores by depression status as defined by the HAMD-17 in this model. Therefore, we calculated mean EQ-5D utilities for remitters, responders and non-responders as defined above ourselves using data from a trial by Kuyken *et al.* (2008) on patients with recurrent depression. To account for the fact that repeated measures were available for most patients, we used a pooled ordinary least-squares model with cluster robust standard errors. We assigned the same quality of life to patients dropping out of treatment due to side effects or no response as to those who completed the treatment but who did not respond.

Costs

We measured costs from a UK health care perspective and used a price year of 2012. Due to lack of robust empirical evidence, we based our costing for the interventions on assumptions made in a previous cost-effectiveness analysis by Simon *et al.* (2006) using unit cost data from Curtis (2012). Since it was the most widely prescribed antidepressant in England in 2010, pharmacotherapy was assumed to consist of a 20 mg daily dose of citalopram over a total of 15 months (Ilyas & Moncrieff, 2012). This is longer than the treatment period suggested by National Institute of Health and Care Excellence (2009) guidance but consistent with the RCTs informing our model. As part of patient monitoring beyond what would be expected in usual care, patients treated with antidepressants were initially assumed to have two appointments with a psychiatric consultant and two with a specialist registrar each lasting 50 min (National Institute of Health and Care Excellence, 2009). CBT treatment was assumed to consist of 16 sessions during the acute treatment phase and two additional 'booster' sessions after that. We assumed that patients who discontinued pharmacotherapy dropped out of treatment after 1 month of treatment and one appointment with a psychiatric consultant whereas patients receiving CBT were assumed to drop out after four sessions. The cost for combination therapy in our model was the sum of the cost of pharmacotherapy and psychotherapy. We assumed that patients who did not respond to the treatment administered in the acute phase would not receive any booster CBT sessions and/or maintenance pharmacotherapy.

We obtained estimates of health care resource use by depression status from the same study that provided the EQ-5D data (Kuyken *et al.* 2008) and again used a pooled ordinary least-squares model with cluster robust standard errors in our estimation of these figures. We updated the costs from this study using the Hospital and Community Health Service Pay and Price Index (Curtis, 2012). To reflect the current value of the benefits and costs accumulating over the time horizon of the model we discounted both at a rate of 3.5% as suggested by the National Institute of Health and Care Excellence (2013). Table 1 summarizes the model inputs that were not estimated in the meta-analyses.

Cost-effectiveness and sensitivity analyses

We calculated incremental cost-effectiveness ratios (ICERs) by dividing the estimated mean differences in costs between two treatments by the mean difference in QALYs. To address uncertainty in the ICERs we undertook a probabilistic sensitivity analysis. This involved repeatedly simulating random draws from the distribution of the parameter inputs in order to determine the joint distribution of the outputs of the model (i.e. the relative mean cost and effects of the interventions). We displayed this distribution on a cost-effectiveness plane as a credible ellipse. This region indicates where the 'true' cost-effectiveness estimate is likely to lie in and can thus be considered to be a two-dimensional generalization of credible intervals.

Given the replications generated by the simulations, it was also possible to determine the net benefit (NB) of each intervention in each of the replications using the formula $NB = \lambda \times E - C$, where C are the costs of the intervention, E the benefits of the intervention in QALYs and λ the value placed on a QALY by decision makers. We then determined the proportion of replications where each of the interventions had the highest net benefit (i.e. the probability that they were the most cost-effective). We displayed these data for a range of λ using cost-effectiveness acceptability curves (CEACs) (Fenwick & Byford, 2005). Values of the CEAC close to 1 or 0 indicated that the uncertainty as to whether the respective treatment was most likely to be the most cost-effective was low (Baio, 2012). In this study, we discuss the results for a range of λ between £20 000 and £30 000 because this has been presumed to be the range of willingness to pay for QALY improvement by NICE but we acknowledge that lower estimates have recently been suggested (Haycox, 2013).

We undertook four sensitivity analyses that had the potential to affect the results of the model. First, we considered the impact of relaxing the inclusion criteria

Table 1. *Input parameters utilized in the decision model (in addition to estimates from the meta-analyses)*

| Parameter	Mean (s.e.)	Source
EQ-5D utility weights		
Remission (HAMD ⩽ 7)	0.80 (0.02)	Kuyken *et al.* (2008)
Response (7 < HAMD < 15)	0.62 (0.04)	
No response (HAMD ⩾ 15)	0.48 (0.05)	
SF-6D utility weights		
Remission (HAMD ⩽ 7)	0.78 (0.01)	Kendrick *et al.* (2009)
Response (7 < HAMD < 15)	0.63 (0.01)	
No response (HAMD ⩾ 15)	0.57 (0.01)	
Cost of interventions, £		
Pharmacotherapy		Assumptions based on Simon *et al.* (2006),
Maintenance treatment	192	Joint Formularly Committee (2013), Curtis (2012)
Acute-phase treatment	676	
Incomplete treatment	282	
CBT		
Booster sessions	208	
Acute-phase treatment	1664	
Incomplete treatment	416	
Monthly health care service use, £		
Remission (HAMD ⩽ 7)	134 (16)	Kuyken *et al.* (2008)
Response (7 < HAMD < 15)	191 (31)	
No response (HAMD ⩾ 15)	352 (89)	
Event probabilities for patients who discontinue treatment		
Probability of feeling cured	0.2 (0.03)	National Institute of Health and Care Excellence (2009);
		Sado *et al.* (2009); Radhakrishnan *et al.* (2013)
Probability of relapsing among those feeling cured	0.8 (0.2)	Jarrett *et al.* (2000)
Probability of relapse among patients responding to treatment		
Pharmacotherapy (absolute risk)	0.78 (0.11)	Kuyken *et al.* (2008)
Annual discount rate (costs and benefits), %	3.5	National Institute of Health and Care Excellence (2013)

s.e., Standard error; EQ-5D, EuroQol 5 Dimensions; HAMD, Hamilton Depression Scale; SF-6D, six-dimensional health state classification from short form-36 (SF-36); CBT, cognitive–behavioural therapy.

for studies to be included in the meta-analysis of acute-phase data. Unlike previous economic models, we focused exclusively on trials assessing CBT or CT. For that reason, amongst others, in our base case meta-analysis we excluded a large trial by Keller *et al.* (2000), which assesses the efficacy of the Cognitive Behavioural-Analysis System of Psychotherapy (CBASP). CBASP shares some of the features of CBT but differs from the way that it is commonly implemented due to its primary focus on interpersonal interactions (Driessen & Hollon, 2010). However, due to the size of this study, in our first sensitivity analysis we analysed the effect of including Keller *et al.* (2000). Second, the costing assumptions with respect to CBT in our base case scenario were likely to favour pharmacotherapy because typically patients tend to attend fewer appointments than set out in this treatment protocol even if completing treatment. In the studies included in the meta-analysis attendance rate was approximately 80%. It was unclear what fraction of unattended sessions was cancelled without prior notice but to gain some understanding of the potential implications of non-attendance we reduced psychotherapy costs by 20% in our second sensitivity analysis. Third, we relaxed the inclusion criteria for the follow-up period, incorporating studies in our meta-analysis in which drug treatment was provided after the end of the acute phase in the pharmacotherapy arm but not necessarily any maintenance treatment in the other two arms. Finally, we used SF-6D (Short Form - six dimensions) utility data from Kendrick *et al.* (2009) to assess the sensitivity of the estimated QALY gains to the choice of utility measure. As in the case of EQ-5D, we used a pooled ordinary least-squares model with robust standard errors to estimate the mean utility scores.

Results

We identified 15 randomized trials that fulfilled our inclusion criteria for the acute phase. Of these, 11 compared pharmacotherapy with CBT as a monotherapy,

Distribution of patients at 3 months

Probability of relapse among remitters
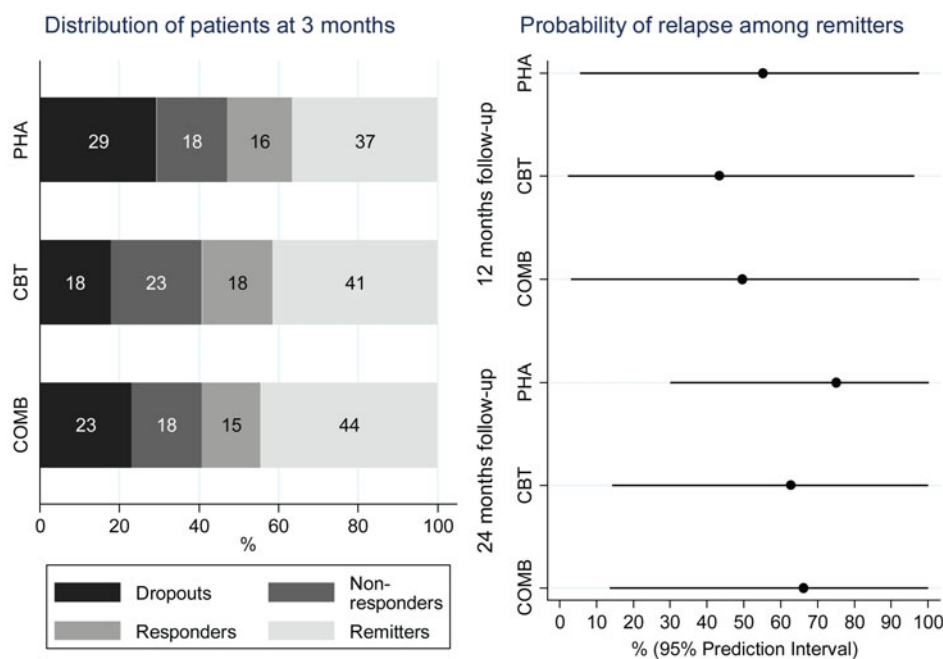
PHA: pharmacotherapy; CBT: cognitive behavioural therapy; COMB: combination therapy

Fig. 2. Results of the meta-analyses.

three compared pharmacotherapy with both CBT and combination treatment, and one compared CBT with combination treatment only. We report the raw data extracted from these studies in online Supplementary Appendix S3. Fig. 2 summarizes the results of the meta-analysis at the end of the acute treatment phase and the meta-analyses at the two follow-ups. CBT and combination therapy both had a higher proportion of patients remitting and a lower proportion of patients discontinuing prematurely than pharmacotherapy. Patients were more likely to complete CBT treatment than combination therapy, but the share of individuals whose depression status improved was similar. The probability of remission was higher under combination treatment than CBT. In the base case scenario, only three trials fulfilled our inclusion criteria for the follow-up phase, one of which included combination therapy. At 12-month follow-up the risk of relapse among remitters was estimated to be lowest for CBT (43%), followed by combination therapy (49%) and pharmacotherapy (55%). The ranking of cumulative relapse rates was the same at 24-month follow-up with probabilities of 62, 66 and 75%, respectively (see Fig. 2). Forest plots for all meta-analyses can be found in online Supplementary Appendix S2.

The model suggested that pharmacotherapy and combination therapy had the lowest and highest expected costs, respectively (£3645 v. £5060). The estimate for CBT was £4418. QALYs were lowest in the pharmacotherapy arm (1.236), whereas those for CBT

and combination treatment were identical up to a rounding error (1.274). Therefore, CBT yielded the same benefits as combination therapy while being less expensive, i.e. CBT dominated combination therapy. The ICER for CBT compared with pharmacotherapy was £20 039/QALY (Table 2). Fig. 3 shows CEACs comparing all three treatments at the same time. Within the aforementioned NICE threshold range there was much uncertainty. At the lower end, pharmacotherapy, the least expensive treatment, was most likely to be cost-effective, whereas at a threshold above around £22 000 per QALY this changes to CBT. There was considerable decision uncertainty because there was a strong chance of not choosing the most cost-effective treatment option when deciding on the treatment with the highest average net benefit, i.e. pharmacotherapy at a willingness to pay of less than £20 039 and CBT above that. Fig. 3 also shows that combination treatment was least likely to be cost-effective with the range of £20 000 to £30 000 per QALY. The estimated probability ranged from 15 to 23%.

Table 2 shows the results of the sensitivity analyses. As previously mentioned, to simplify comparison between the scenarios, we also display the estimated cost and effect differences for each of these graphically using cost-effectiveness planes in online Supplementary Appendix S1. Including the study by Keller *et al.* (2000) in the meta-analysis substantially increased the cost-effectiveness of combination treatment, such that

Table 2. *Cost-effectiveness results for the base case and sensitivity analyses*

	Intervention	Difference in costs compared with PHA, £ (95% CrI)	Difference in QALYs to PHA (95% CrI)	ICER, £/QALY	Probability of being the most cost-effective treatment at a WTP of £25 000/QALY, %
Base case analysis	CBT	773 (470 to 1036)	0.038 (−0.032 to 0.128)	20 039	43
	COMB	1415 (943 to 1802)	0.038 (−0.048 to 0.151)	N.A.	20
Sensitivity analysis I: including Keller *et al.* (2000)	CBT	766 (465 to 1043)	0.036 (−0.035 to 0.132)	21 448	31
	COMB	1385 (953 to 1772)	0.062 (−0.025 to 0.178)	23 807	37
Sensitivity analysis II: 80% CBT attendance rate	CBT	458 (201 to 703)	0.038 (−0.032 to 0.128)	11 922	51
	COMB	1112 (727 to 1462)	0.038 (−0.048 to 0.151)	740 776	23
Sensitivity analysis III: relaxed inclusion criteria for follow-up	CBT	777 (488 to 1031)	0.038 (−0.031 to 0.125)	20 534	39
	COMB	1387 (941 to 1760)	0.048 (−0.042 to 0.156)	61 403	28
Sensitivity analysis IV: QALYs estimated using SF-6D	CBT	774 (475 to 1033)	0.024 (−0.022 to 0.078)	32 582	32
	COMB	1416 (953 to 1801)	0.025 (−0.03 to 0.094)	386 154	19

PHA, Pharmacotherapy; CrI, credible interval; QALY, quality-adjusted life year; ICER, incremental cost-effectiveness ratio; WTP, willingness to pay; CBT, cognitive–behavioural therapy/cognitive therapy; COMB, combination therapy; N.A., not applicable (intervention is dominated); SF-6D, six-dimensional health state classification from short form-36 (SF-36).
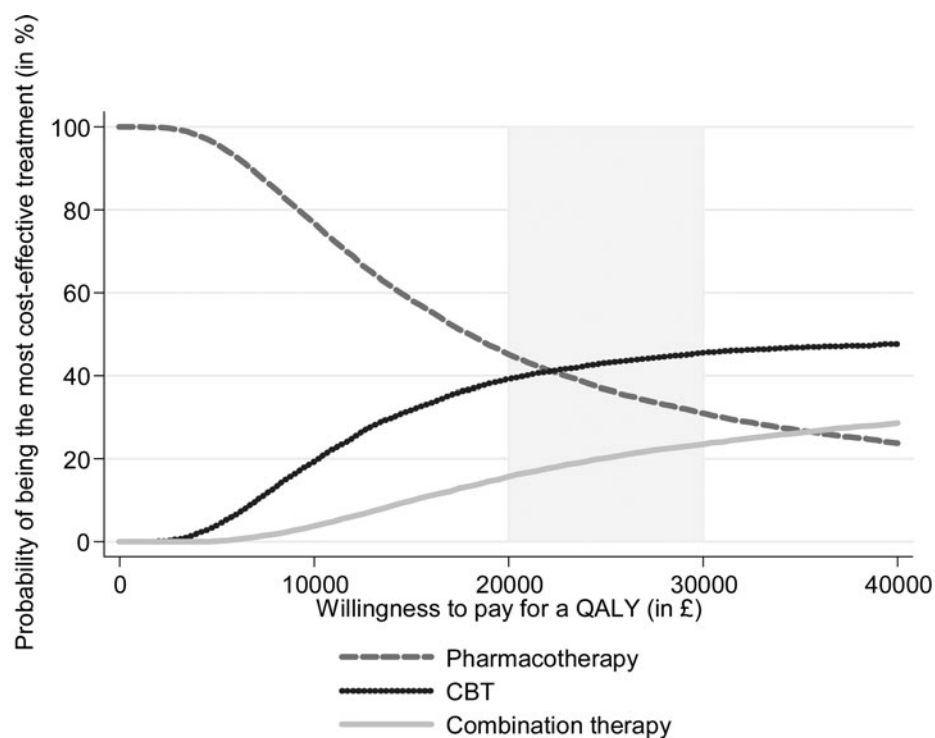
Fig. 3. Cost-effectiveness acceptability curves. QALY, Quality-adjusted life year; CBT, cognitive–behavioural therapy.

it became the treatment most rather than least likely to be cost-effective. However, the decision uncertainty between the three treatment options remained high in this scenario. As indicated in Table 2 at a willingness to pay of £25 000 per QALY, the probabilities of pharmacotherapy (32%), CBT (31%) and combination treatment (37%) being the most cost-effective treatment were comparable. As expected, reducing the CBT attendance rate in the second sensitivity analysis lowered the costs of both CBT monotherapy and combination treatment. This did not affect the ranking of treatments in terms of their cost-effectiveness but the ICER for CBT was considerably lower at £9714 per QALY in this scenario. In the third sensitivity analysis, two additional studies were included in the meta-analysis of relapse rates. As a result, rather than CBT, relapse rates were estimated to be lowest with combination therapy. However, the additional QALY gain of combination treatment over CBT was not sufficient to make it the treatment with the highest likelihood of being cost-effective and at £61 403 per QALY the ICER for combination therapy relative to CBT was high. Using SF-6D instead of EQ-5D to estimate QALYs decreased the absolute differences in QALYs between the treatments such that the ICER for CBT increased to £32 582 per QALY and at a willingness to pay for a QALY of £25 000 pharmacotherapy became the treatment most likely to be cost-effective (51% probability). In summary, two sensitivity analyses affected the ranking of the interventions in terms of the probability of being the most cost-effective compared with our base case scenario and the ICERs for the interventions varied markedly between the scenarios under consideration.

Discussion

This study adds to the existing economic evaluations comparing psychological and pharmacological treatments for depression. Specifically, two decision models commissioned by NICE exist examining a comparable decision problem as our study. However, these analyses come to conclusions different from our appraisal of the evidence. The National Institute of Health and Care Excellence (2009) reported ICERs of £7052 and £5558 per QALY, respectively, for combination therapy relative to pharmacotherapy in the treatment of moderate and severe depression in secondary care. In the related study (Simon *et al.* 2006) the corresponding ICERs were £4056 and £14 540 per QALY. By contrast, our analysis suggested that CBT as monotherapy was more likely to be cost-effective than pharmacotherapy which in turn was more likely to be cost-effective than combination treatment. In addition, the decision uncertainty was much greater in our model, i.e. the

difference in the probability of being cost-effective between the treatments as suggested by the CEACs was lower.

These discrepancies are largely due to four differences in the modelling approach. First, we specified inclusion criteria different from these earlier studies. For example, we excluded data by Simons *et al.* (1986) which suggested a large difference in relapse rates between pharmacotherapy and combination therapy because, contrary to NICE guidance, treatment was discontinued in this study following the end of the acute phase. Also, we restricted our analysis to studies on CBT or CT rather than allowing for all types of psychotherapies. In our sensitivity analysis we showed that excluding the trial by Keller *et al.* (2000), which reported large benefits of combination treatment over the two monotherapies for a type of psychological therapy that is not commonly provided in the NHS, has a significant effect on the cost-effectiveness estimate of combination treatment. Second, we used more modern methods of evidence synthesis. As a result, for example, when comparing the effect of pharmacotherapy and combination therapy on remission rates, the network meta-analysis estimates are smaller than those in a traditional pairwise analysis because network meta-analysis accounts for the fact that in trials including all three treatments, the treatment effect for CBT was above average compared with the trials which did not include combination treatment. In statistical terms, network meta-analysis accounted for trial arms not missing completely at random. Third, the time horizon of the study was extended beyond 15 months. This particularly improved the cost-effectiveness of CBT since, at least in the base case scenario, the meta-analysis suggested that was superior to the other two treatments in preventing relapse. Fourth, our model was based on different health states and disease trajectories. For example, we differentiated between patients showing full and partial response rather than implicitly assuming that all patients who did not remit showed no response. On the other hand, we considered the information reported in clinical trials to be too limited to differentiate between the treatment effects for patients with moderate depression and those with severe depression at baseline. Instead, we modelled the cost-effectiveness for these two populations as a whole.

Besides these differences in the cost-effectiveness estimates, we would also argue that the conclusions that can be drawn from the results of our economic model and its predecessors are more limited than implied by National Institute of Health and Care Excellence (2009) guidelines. Current recommendations do not differentiate between treatment in primary and secondary care settings but the vast majority of the

data used to inform the parameters in this model as well as the analyses by the National Institute of Health and Care Excellence (2009) and Simon *et al.* (2006) draw from the clinical literature set in secondary care. One would expect depression severity and general health status of patients to be worse in this setting and there are also likely to be different implications for resource use than in primary care. Therefore, we would argue that the conclusions drawn from this model cannot easily be transferred to treatment of depression in primary care, the setting in which most patients with depression are cared for in the UK. The differences between the results of this model and its predecessors do, however, illustrate the fragility of cost-effectiveness estimates. Given investments of over £400 million to increase the availability of psychological treatments in primary care as part of the Improving Access to Psychological Therapies (IAPT) initiative, this also highlights the need for further examination of the economic evidence for treatment of depression in UK primary care (Clark, 2011; Department of Health, 2012).

We invite the reader to critically examine our choice of inputs, synthesis method, model structure and the validity of the conclusions drawn in light of the available evidence (Afzali *et al.* 2012*a*, *b*). Decision models can be useful in synthesizing, linking and extrapolating a wide variety of information in a transparent fashion, but they are inevitably reductionist and can only be regarded as decision aids (something that could be also argued to apply to RCTs). In the appraisal of the evidence base, modellers are faced with an inevitable trade-off between competing objectives such as precision, relevance, validity and feasibility. This model is no exception from most other health economic models in that it is heavily dependent on structural assumptions, i.e. it largely omits non-statistical uncertainty (Grutters *et al.* 2015).

Our meta-analyses combined heterogeneous treatments to yield a single treatment effect estimate. Thus, rather than taking the 'best' antidepressant as the relevant benchmark, as was the case in analyses by the National Institute of Health and Care Excellence (2009) and Simon *et al.* (2006), we evaluated pharmacotherapy and combination therapy as a class which may have biased results to some extent. Similarly, both in psychological therapy and pharmacotherapy, variations in non-specific elements, treatment fidelity and the therapeutic relationship across studies are likely to have some bearing on treatment effectiveness which may have been a source of bias. However, there was a dearth of evidence with respect to some model parameters, particularly response and relapse rates in combination therapy. Therefore, we believe that our approach was preferable to applying of

further quality criteria as this would have further reduced the robustness of the model or required additional subjective judgments. Unlike previous studies, we consistently used the HAMD as the backbone of the model to enhance the internal consistency of the model; however, to do so it was necessary to derive key model parameter estimates from a single trial carried out in a primary care setting (Kuyken *et al.* 2008). Since patients in primary care are likely to consume fewer health care resources beyond the resources used for depression treatment, this is likely to have underestimated the cost-effectiveness of the more effective treatments, i.e. combination treatment and CBT, to some extent. Furthermore, the psychometric properties of the HAMD have been debated and may have biased our comparison (Fountoulakis *et al.* 2014). This study applies more modern statistical methods to estimate parameters and account for uncertainty than its predecessors, but, amongst others, a limitation which remains in our meta-analysis is the assumption that there is no correlation in outcomes within treatment arms which is unlikely to hold.

Another feature of our model that should be emphasized is that, like its precursors, the efficacy of the interventions was based primarily on RCT data given its status in the NICE hierarchy of evidence and because adequate observational data were not available. Yet, this also represents a limitation of this study because the generalizability of RCTs in depression to routine clinical practice has been questioned (van der Lem *et al.* 2012). Since reporting on side effects, suicide risks and the exact timing of events in these trials is inconsistent or incomplete, this also meant that we did not incorporate these into the decision model. The exclusion of adverse drug reactions evidently favoured pharmacotherapy and combination therapy but there does not appear to be clear evidence to indicate what the direction of bias in our model might be because of the omission of the other two factors. In addition, in all relevant RCTs antidepressants were tapered off after 6 or 12 months of follow-up for all patients in the pharmacotherapy and combination treatment arms. However, due to increased risk of relapse after withdrawal from medications, National Institute of Health and Care Excellence (2009) guidelines advise continuation of antidepressant treatments for longer than that for at-risk patients which is increasingly the case in routine clinical practice (Moore *et al.* 2009). It is unclear what effect this discrepancy between naturalistic and trial prescribing practice might have on the relative cost-effectiveness of pharmacotherapy and combination treatment. We would also like to emphasize that, at best, data from RCTs only apply to individuals who are likely to participate in them. Some

patients treated for depression in secondary care are likely to be so unwell that they are unlikely to be enrolled in a typical trial and in practice many patients with MDD have strong preferences for pharmacotherapy or psychotherapy which should be taken into account in the choice of treatment for depression (van Schaik *et al.* 2004; National Institute of Health and Care Excellence, 2009).

Conclusion

In this paper we have revisited the evidence base for treatment of depression in UK secondary care. Our economic appraisal differs from the recommendation made by current National Institute of Health and Care Excellence (2009) guidance and our decision model may therefore be of use when reviewing treatment recommendations for depression in the future. In a routine care setting, local consideration and the constraints in the supply of CBT are likely to play a role in treatment decisions, but this model may offer a platform for further discussion of the cost-effectiveness of interventions (Gyani *et al.* 2012). We found the evidence base comparing pharmacotherapy, CBT and combination therapy for depression to be remarkably limited, particularly with respect to the latter. Thus, besides further economic evaluations in UK primary care, we would like to suggest three areas of uncertainty that, we believe, would particularly warrant more in-depth exploration. First, the relative cost-effectiveness of different types of interventions for depression is likely to differ between subgroups of patients. Personalizing the choice of treatment will be critical to more efficient use of health care resources in the future (Simon & Perlis, 2010; Cuijpers *et al.* 2012; Wallace *et al.* 2013; Hollinghurst *et al.* 2014). Second, we would encourage the use of observational data coupled with appropriate statistical methods to gain a better understanding of the effects of depression treatment in a routine care setting and the long-term cost-effectiveness. Third, the wider societal impact of depression treatments and the way that they should be incorporated into health technology assessment are currently unclear due to a host of methodological, ethical, policy and practical issues (Sculpher, 2001; National Institute of Health and Care Excellence, 2014). However, recent evidence suggests that CBT may produce greater improvements in employability which warrants further exploration (Fournier *et al.* 2015).

Supplementary material

For supplementary material accompanying this paper visit http://dx.doi.org/10.1017/S0033291715000951

Acknowledgements

Declaration of interest

None

References

Afzali HHA, Karnon J, Gray J (2012*a*). A critical review of model-based economic studies of depression. *Pharmacoeconomics* **30**, 461–482.

Afzali HHA, Karnon J, Gray J (2012*b*). A proposed model for economic evaluations of major depressive disorder. *European Journal of Health Economics* **13**, 501–510.

American Psychiatric Association (2000). *Handbook of Psychiatric Measures*. American Psychiatric Publishing, Inc.: Washington, DC.

Baio G (2012). *Bayesian Methods in Health Economics*. Chapman and Hall/CRC Press: Boca Raton, FL.

Borenstein M, Hedges LV, Higgins JPT, Rothstein HR (2009). *Introduction to Meta-Analysis*. Wiley: Chichester.

Clark DM (2011). Implementing NICE guidelines for the psychological treatment of depression and anxiety disorders: the IAPT experience. *International Review of Psychiatry* **23**, 318–327.

Cuijpers P, Reynolds CF, Donker T, Li J, Andersson G, Beekman A (2012). Personalized treatment of adult depression: medication, psychotherapy, or both? A systematic review. *Depression and Anxiety* **29**, 855–864.

Cuijpers P, van Straten A, Warmerdam L, Andersson G (2008). Psychological treatment of depression: a meta-analytic database of randomized studies. *BMC Psychiatry* **8**, 36.

Curtis L (2012). *Unit Costs of Health and Social Care 2012*. Personal Social Services Research Unit: Canterbury.

Department of Health (2012). IAPT three-year report – the first million patients. Department of Health: London.

Dias S, Sutton AJ, Ades AE, Welton NJ (2013*a*). Evidence synthesis for decision making 2: a generalized linear modeling framework for pairwise and network meta-analysis of randomized controlled trials. *Medical Decision Making* **33**, 607–617.

Dias S, Sutton AJ, Welton NJ, Ades AE (2013*b*). Evidence synthesis for decision making 3: heterogeneity – subgroups, meta-regression, bias, and bias-adjustment. *Medical Decision Making* **33**, 618–640.

Dias S, Welton NJ, Sutton AJ, Ades AE (2013*c*). Evidence synthesis for decision making 5: the baseline natural history model. *Medical Decision Making* **33**, 657–670.

Driessen E, Hollon SD (2010). Cognitive behavioral therapy for mood disorders: efficacy, moderators and mediators. *Psychiatric Clinics of North America* **33**, 537–555.

Fenwick E, Byford S (2005). A guide to cost-effectiveness acceptability curves. *British Journal of Psychiatry* **187**, 106–108.

Fountoulakis KN, Samara MT, Siamouli M (2014). Burning issues in the meta-analysis of pharmaceutical trials for depression. *Journal of Psychopharmacology* **28**, 106–117.

Fournier JC, DeRubeis RJ, Amsterdam J, Shelton RC, Hollon SD (2015). Gains in employment status following antidepressant medication or cognitive therapy for depression. *British Journal of Psychiatry* **206**, 332–338.

Gelman A, Jakulin A, Pittau MG, Su Y-S (2008). A weakly informative default prior distribution for logistic and other regression models. *Annals of Applied Statistics* **2**, 1360–1383.

Grutters JC, van Asselt MA, Chalkidou K, Joore M (2015). Healthy decisions: towards uncertainty tolerance in healthcare policy. *Pharmacoeconomics* **33**, 1–4.

Gyani A, Pumphrey N, Parker H, Shafran R, Rose S (2012). Investigating the use of NICE guidelines and IAPT services in the treatment of depression. *Mental Health in Family Medicine* **9**, 149–160.

Haycox A (2013). How much should the NHS pay for a QALY? *Pharmacoeconomics* **31**, 357–359.

Higgins JP, Green S (2008). *Cochrane Handbook for Systematic Reviews of Interventions*. Wiley Online Library: Chichester.

Hollinghurst S, Carroll FE, Abel A, Campbell J, Garland A, Jerrom B, Kessler D, Kuyken W, Morrison J, Ridgway N, Thomas L, Turner K, Williams C, Peters TJ, Lewis G, Wiles N (2014). Cost-effectiveness of cognitive–behavioural therapy as an adjunct to pharmacotherapy for treatment-resistant depression in primary care: economic evaluation of the CoBalT Trial. *British Journal of Psychiatry* **204**, 69–76.

Ilyas S, Moncrieff J (2012). Trends in prescriptions and costs of drugs for mental disorders in England, 1998–2010. *British Journal of Psychiatry* **200**, 393–398.

Jarrett RB, Kraft D, Schaffer M, Witt-Browder A, Risser R, Atkins DH, Doyle J (2000). Reducing relapse in depressed outpatients with atypical features: a pilot study. *Psychother Psychosom* **69**, 232–239.

Joint Formulary Committee (2013). *British National Formulary*, 66th edn. BMJ Group and Pharmaceutical Press: London.

Keller MB, McCullough JP, Klein DN, Arnow B, Dunner DL, Gelenberg AJ, Markowitz JC, Nemeroff CB, Russell JM, Thase ME, Trivedi MH, Blalock JA, Borian FE, Jody DN, DeBattista C, Koran LM, Schatzberg AF, Fawcett J, Hirschfeld RMA, Keitner G, Miller I, Kocsis JH, Kornstein SG, Manber R, Ninan PT, Rothbaum B, Rush AJ, Vivian D, Zajecka J (2000). A comparison of nefazodone, the cognitive behavioral–analysis system of psychotherapy, and their combination for the treatment of chronic depression. *New England Journal of Medicine* **342**, 1462–1470.

Kendrick T, Chatwin J, Dowrick C, Tylee A, Morriss R, Peveler R, Leese M, McCrone P, Harris T, Moore M, Byng R, Brown G, Barthel S, Mander H, Ring A, Kelly V, Wallace V, Gabbay M, Craig T, Mann A (2009). Randomised controlled trial to determine the clinical effectiveness and cost-effectiveness of selective serotonin reuptake inhibitors plus supportive care, *versus* supportive care alone, for mild to moderate depression with somatic symptoms in primary care: the THREAD (THREshold for AntiDepressant response) study. *Health Technol Assess* **13**, iii–iv, ix-xi, 1–159.

Kuyken W, Byford S, Taylor RS, Watkins E, Holden E, White K, Barrett B, Byng R, Evans A, Mullan E, Teasdale JD (2008). Mindfulness-based cognitive therapy to prevent relapse in recurrent depression. *J Consult Clin Psychol* **76**, 966–978.

Malek MH (2000). *Implementing QALYs*. Hayward Medical Communications: London.

McCrone P, Dhanasiri S, Patel A, Knapp M, Lawton-Smith S (2008). *Paying the Price: the Cost of Mental Health Care in England to 2026*. King's Fund: London.

Moore M, Yuen HM, Dunn N, Mullee MA, Maskell J, Kendrick T (2009). Explaining the rise in antidepressant prescribing: a descriptive study using the general practice research database. *BMJ* **339**, b3999.

National Institute of Health and Care Excellence (2009). Depression: the NICE guideline on the treatment and management of depression – updated edition. In *National Clinical Practice Guideline 90* (ed. National Collaborating Centre for Mental Health). The British Psychological Society and The Royal College of Psychiatrists: London.

National Institute of Health and Care Excellence (2013). *Guide to the Methods of Technology Appraisal 2013*. National Institute for Health and Care Excellence: London.

National Institute of Health and Care Excellence (2014). *Value Based Assessment of Health Technologies*. National Institute for Health and Care Excellence: London.

Radhakrishnan M, Hammond G, Jones PB, Watson A, McMillan-Shields F, Lafortune L (2013). Cost of Improving Access to Psychological Therapies (IAPT) programme: an analysis of cost of session, treatment and recovery in selected primary care trusts in the East of England region. *Behaviour Research and Therapy* **51**, 37–45.

Riley RD, Higgins JPT, Deeks JJ (2011). Interpretation of random effects meta-analyses. *BMJ* **342**, d549.

Sado M, Knapp M, Yamauchi K, Fujisawa D, So M, Nakagawa A, Kikuchi T, Ono Y (2009). Cost-effectiveness of combination therapy *versus* antidepressant therapy for management of depression in Japan. *Australian and New Zealand Journal of Psychiatry* **43**, 539–547.

Sculpher M (2001). The role and estimation of productivity costs in economic evaluation. In *Economic Evaluation in Health Care: Merging Theory with Practice* (ed. M Drummond and A McGuire), pp. 94–112. Oxford University Press: Oxford.

Simon GE, Perlis RH (2010). Personalized medicine for depression: can we match patients with treatments? *American Journal of Psychiatry* **167**, 1445–1455.

Simon J, Pilling S, Burbeck R, Goldberg D (2006). Treatment options in moderate and severe depression: decision analysis supporting a clinical guideline. *British Journal of Psychiatry* **189**, 494–501.

Simons AD, Murphy GE, Levine JL, Wetzel RD (1986). Cognitive therapy and pharmacotherapy for depression.

Sustained improvement over one year. *Archives of General Psychiatry* **43**, 43–48.

Turner RM, Davey J, Clarke MJ, Thompson SG, Higgins JP (2012). Predicting the extent of heterogeneity in meta-analysis, using empirical data from the Cochrane Database of Systematic Reviews. *International Journal of Epidemiology* **41**, 818–827.

van der Lem R, van der Wee NJA, van Veen T, Zitman FG (2012). Efficacy *versus* effectiveness: a direct comparison of the outcome of treatment for mild to moderate depression in randomized controlled trials and daily practice. *Psychotherapy and Psychosomatics* **81**, 226–234.

van Schaik DJF, Klijn AFJ, van Hout HPJ, van Marwijk HWJ, Beekman ATF, de Haan M, van Dyck R (2004). Patients' preferences in the treatment of depressive disorder in primary care. *General Hospital Psychiatry* **26**, 184–189.

Wallace ML, Frank E, Kraemer HC (2013). A novel approach for developing and interpreting treatment moderator profiles in randomized clinical trials. *JAMA Psychiatry* **70**, 1241–1247.