

Lithium or an atypical antipsychotic drug in the management of treatment-resistant depression: a systematic review and economic evaluation

SJ Edwards, V Hamilton, L Nherera and N Trevor



***National Institute for
Health Research***

Lithium or an atypical antipsychotic drug in the management of treatment-resistant depression: a systematic review and economic evaluation

SJ Edwards,* V Hamilton, L Nherera and N Trevor

BMJ Technology Assessment Group (BMJ-TAG), London, UK

*Corresponding author

Declared competing interests of authors: SJE was an employee of AstraZeneca UK Ltd until August 2010. AstraZeneca holds the marketing authorisation for Seroquel® (quetiapine), an atypical antipsychotic drug that is included in this report.

Published November 2013

DOI: 10.3310/hta17540

This report should be referenced as follows:

Edwards SJ, Hamilton V, Nherera L, Trevor N. Lithium or an atypical antipsychotic drug in the management of treatment-resistant depression: a systematic review and economic evaluation. *Health Technol Assess* 2013;**17**(54).

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

ISSN 1366-5278 (Print)

ISSN 2046-4924 (Online)

Five-year impact factor: 5.804

Health Technology Assessment is indexed in MEDLINE, CINAHL, EMBASE, The Cochrane Library and the ISI Science Citation Index and is assessed for inclusion in the Database of Abstracts of Reviews of Effects.

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

Editorial contact: nihredit@southampton.ac.uk

The full HTA archive is freely available to view online at www.journalslibrary.nihr.ac.uk/hta. Print-on-demand copies can be purchased from the report pages of the NIHR Journals Library website: www.journalslibrary.nihr.ac.uk

Criteria for inclusion in the *Health Technology Assessment* journal

Reports are published in *Health Technology Assessment* (HTA) if (1) they have resulted from work for the HTA programme, and (2) they are of a sufficiently high scientific quality as assessed by the reviewers and editors.

Reviews in *Health Technology Assessment* are termed 'systematic' when the account of the search appraisal and synthesis methods (to minimise biases and random errors) would, in theory, permit the replication of the review by others.

HTA programme

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

The journal is indexed in NHS Evidence via its abstracts included in MEDLINE and its Technology Assessment Reports inform National Institute for Health and Care Excellence (NICE) guidance. HTA research is also an important source of evidence for National Screening Committee (NSC) policy decisions.

For more information about the HTA programme please visit the website: www.hta.ac.uk/

This report

The research reported in this issue of the journal was funded by the HTA programme as project number 10/30/01. The contractual start date was in July 2011. The draft report began editorial review in June 2012 and was accepted for publication in October 2012. The authors have been wholly responsible for all data collection, analysis and interpretation, and for writing up their work. The HTA editors and publisher have tried to ensure the accuracy of the authors' report and would like to thank the reviewers for their constructive comments on the draft document. However, they do not accept liability for damages or losses arising from material published in this report.

This report presents independent research funded by the National Institute for Health Research (NIHR). The views and opinions expressed by authors in this publication are those of the authors and do not necessarily reflect those of the NHS, the NIHR, NETSCC, the HTA programme or the Department of Health. If there are verbatim quotations included in this publication the views and opinions expressed by the interviewees are those of the interviewees and do not necessarily reflect those of the authors, those of the NHS, the NIHR, NETSCC, the HTA programme or the Department of Health.

© Queen's Printer and Controller of HMSO 2013. This work was produced by Edwards *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

Published by the NIHR Journals Library (www.journalslibrary.nihr.ac.uk), produced by Prepress Projects Ltd, Perth, Scotland (www.prepress-projects.co.uk).

Editor-in-Chief of *Health Technology Assessment* and NIHR Journals Library

Professor Tom Walley Director, NIHR Evaluation, Trials and Studies and Director of the HTA Programme, UK

NIHR Journals Library Editors

Professor Ken Stein Chair of HTA Editorial Board and Professor of Public Health, University of Exeter Medical School, UK

Professor Andree Le May Chair of NIHR Journals Library Editorial Group (EME, HS&DR, PGfAR, PHR journals)

Dr Martin Ashton-Key Consultant in Public Health Medicine/Consultant Advisor, NETSCC, UK

Professor Matthias Beck Chair in Public Sector Management and Subject Leader (Management Group), Queen's University Management School, Queen's University Belfast, UK

Professor Aileen Clarke Professor of Health Sciences, Warwick Medical School, University of Warwick, UK

Dr Tessa Crilly Director, Crystal Blue Consulting Ltd, UK

Dr Peter Davidson Director of NETSCC, HTA, UK

Ms Tara Lamont Scientific Advisor, NETSCC, UK

Professor Elaine McColl Director, Newcastle Clinical Trials Unit, Institute of Health and Society, Newcastle University, UK

Professor William McGuire Professor of Child Health, Hull York Medical School, University of York, UK

Professor Geoffrey Meads Honorary Professor, Business School, Winchester University and Medical School, University of Warwick, UK

Professor Jane Norman Professor of Maternal and Fetal Health, University of Edinburgh, UK

Professor John Powell Consultant Clinical Adviser, National Institute for Health and Care Excellence (NICE), UK

Professor James Raftery Professor of Health Technology Assessment, Wessex Institute, Faculty of Medicine, University of Southampton, UK

Dr Rob Riemsma Reviews Manager, Kleijnen Systematic Reviews Ltd, UK

Professor Helen Roberts Professorial Research Associate, University College London, UK

Professor Helen Snooks Professor of Health Services Research, Institute of Life Science, College of Medicine, Swansea University, UK

Please visit the website for a list of members of the NIHR Journals Library Board:
www.journalslibrary.nihr.ac.uk/about/editors

Editorial contact: nihredit@southampton.ac.uk

Abstract

Lithium or an atypical antipsychotic drug in the management of treatment-resistant depression: a systematic review and economic evaluation

SJ Edwards,* V Hamilton, L Nherera and N Trevor

BMJ Technology Assessment Group (BMJ-TAG), London, UK

*Corresponding author

Background: Patients with treatment-resistant depression (TRD) are those with major depressive disorder that has not responded adequately to treatment. The causes of depression are not fully understood, although there is evidence to suggest that depression is a complex interaction among biological, genetic, psychosocial and environmental factors. Strategies available for the treatment of patients with TRD include pharmacological, non-pharmacological, and psychological and psychosocial interventions. Pharmacological treatment options include switching to a different antidepressant, the addition of another antidepressant of a different class, or use of an augmenting agent, such as anticonvulsants, lithium or atypical antipsychotics (AAPs). However, there is limited evidence available on the effectiveness of these strategies in the treatment of TRD.

Objectives: To estimate the clinical effectiveness and cost-effectiveness of augmentation of selective serotonin reuptake inhibitor (SSRI) antidepressant therapy with either lithium or an AAP drug in the management of people with treatment-resistant unipolar depression, defined as failure to respond to two or more antidepressant drugs in their current episode of depression.

Data sources: Databases searched were Cochrane Central Register of Controlled Trials (CENTRAL), EMBASE, MEDLINE, PsycINFO and NHS Economic Evaluation Database (NHS EED). All databases were searched from inception to August 2011. Additional data were obtained from manufacturers.

Review methods: Systematic reviews of studies evaluating clinical effectiveness, economic analyses and quality of life (QoL) were executed. Quality assessment according to predefined criteria was undertaken independently by two reviewers. Pairwise meta-analyses and mixed-treatment comparisons (MTCs) using both fixed- and random-effects models were undertaken based on intention-to-treat analyses. A probabilistic de novo mathematical model was developed to synthesise the available data on costs and clinical outcomes from the UK NHS perspective over a 1-year time horizon (8 weeks of acute treatment captured by a decision tree and 10 months of maintenance treatment captured by a Markov model).

Results: Twelve randomised controlled trials (RCTs) were identified in the review of clinical effectiveness literature; 10 considered SSRI + AAP compared with SSRI + placebo/no treatment, one considered SSRI + AAP compared with SSRI + lithium and one considered SSRI + lithium compared with SSRI + placebo. The RCTs included in the primary analyses used fluoxetine as the background SSRI and olanzapine as the AAP. Results of the MTC showed a non-significant trend in favour of lithium augmentation for response [lithium a priori odds ratio (OR) 1.29; 95% credible interval (CrI) 0.11 to 5.32; lithium post hoc OR 4.15; 95% CrI 0.25 to 20.34 (the trial informing the comparison with lithium reported response using two different definitions)], mean change in Montgomery–Åsberg Depression Rating Scale score from baseline (mean difference –1.47, 95% CrI –9.10 to 6.41) and all-cause withdrawals (OR 0.74, 95% CrI 0.10 to 2.66). Four economic evaluations (none directly addressing the review question) and 17 studies that reported on

QoL were identified and summarised in narrative reviews. The results of the de novo modelling indicate that augmentation of SSRI with lithium dominates augmentation of an SSRI with AAP (i.e. it resulted in cost savings of £905 per person per year and generated more health benefits, estimated to be 0.03 quality-adjusted life-years). However, sensitivity analyses showed that the model was highly sensitive to changes in acute treatment efficacy (response and remission) or discontinuation. The model was not sensitive to changes in other parameters.

Limitations: In patients with TRD, there is a lack of direct evidence comparing the clinical effectiveness of augmenting an SSRI with an AAP compared with augmenting with lithium. RCTs were identified which facilitated comparison of adding AAP with adding lithium via a MTC. However, variations in the definitions of response implemented in the RCTs, together with differences in patient baseline characteristics across RCTs, introduce bias into the analysis. The direction and extent of the bias is uncertain.

Conclusions: Augmentation of SSRIs with lithium or AAP is likely to be beneficial in people with TRD. Clinical evaluation based on the limited evidence identified in this research indicates no statistically significant difference between the two augmentation strategies. Cost-effectiveness analyses suggest that augmentation with lithium is less expensive and more effective than augmentation with AAP. However, the uncertainty in the clinical estimates of discontinuation and treatment response is reflected in the model results. A RCT comparing the two augmentation strategies, reporting relevant outcomes, including QoL, is needed.

Study registration: PROSPERO CRD42011001464.

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

Contents

Glossary	ix
List of abbreviations	xi
Scientific summary	xiii
Chapter 1 Background	1
Description of health problem	1
Incidence and prevalence	2
Current service provision	3
Current service cost	4
Description of technology under assessment	6
Chapter 2 Definition of the decision problem	11
Decision problem	11
Overall aims and objectives of assessment	12
Chapter 3 Assessment of clinical effectiveness	13
Methods for reviewing effectiveness	13
Results	17
Chapter 4 Assessment of cost-effectiveness	45
Economic literature review	45
Quality-of-life literature review	55
Chapter 5 De novo economic analysis	65
Introduction	65
Model overview	65
Model structure	67
Effectiveness data	68
Costs	74
Sensitivity analysis	81
Results	85
Health economics discussion	92
Chapter 6 Overall discussion	97
Statement of principal findings	97
Summary of economic findings	101
Strengths and limitations and uncertainty of the assessment	101
Chapter 7 Conclusions	105
Implications for service provision	105
Suggested research priorities	105
Acknowledgements	107
References	109

Appendix 1 Final protocol	119
Appendix 2 Literature search strategies for the clinical review	141
Appendix 3 Copy of data extraction form	147
Appendix 4 Table of excluded clinical effectiveness studies with rationale	155
Appendix 5 Quality assessment of clinical trials	159
Appendix 6 Summary of trials included for each outcome in the network meta-analysis	169
Appendix 7 Health economics literature search strategies	171
Appendix 8 Quality assessment of included studies (cost-effectiveness studies)	181
Appendix 9 Table of excluded health economics studies with rationale	187

Glossary

Cost-effectiveness acceptability curves A graph that plots a range of possible cost-effectiveness thresholds on the horizontal axis against the probability (chance) that the intervention will be cost-effective on the vertical axis. In technology appraisals, cost-effectiveness acceptability curves are a means of representing the uncertainty surrounding the cost-effectiveness estimates in relation to the decision.

Cost-effectiveness model An explicit mathematical framework, which is used to represent clinical decision problems and incorporate evidence from a variety of sources to estimate costs and health outcomes.

Discounting Costs and benefits incurred today are usually valued more highly than costs and benefits occurring in the future. Discounting health benefits reflects society's preference for benefits to be experienced in the present rather than the future. Discounting costs reflects society's preference for costs to be experienced in the future rather than the present.

Dominance An intervention is dominated if it has higher costs and worse outcomes than an alternative intervention.

Incremental cost-effectiveness ratio The ratio of the difference in the mean costs of a technology compared with the next best alternative to the differences in the mean outcomes.

Mixed-treatment comparison An analysis that compares two or more interventions using a combination of direct evidence (from head-to-head trials of the interventions of interest) and indirect evidence (trials that do not compare the interventions of interest directly in head-to-head trials but have a common comparator).

QT interval Time interval on an electrocardiogram that represents the interval between the start of the electrical stimulation of the ventricles (the Q wave) and the end of the recharging of the electrical cycle in the heart (the T wave).

Quality-adjusted life-year An index of survival that is adjusted to account for the patient's quality of life during this time. Quality-adjusted life-years have the advantage of incorporating changes in both quantity and quality of life and are used to measure benefits in cost-utility analysis.

List of abbreviations

5-HT	5-hydroxytryptamine	GP	general practitioner
AAP	atypical antipsychotic drug	HAMD	Hamilton Depression Rating Scale
ATHF	antidepressant treatment history form	HCP	health-care professional
BAP	British Association of Psychopharmacology	HRQoL	health-related quality of life
b.i.d.	twice daily	HSUV	health-state utility value
BMI	body mass index	HTA	Health Technology Assessment
BNF	<i>British National Formulary</i>	HUI	Health Utilities Index
CCDAN	Cochrane Collaboration Depression, Anxiety and Neurosis Review Group	ICD-10	<i>International Classification of Diseases, Tenth Revision</i>
CCDANCTR	CCDAN Controlled Trials Register	ICER	incremental cost-effectiveness ratio
CCMD-3	<i>Chinese Classification of Mental Disorders, Version 3</i>	ITT	intention to treat
CEAC	cost-effectiveness acceptability curve	LLFDI	Late-Life Function and Disability Instrument
CI	confidence interval	MADRS	Montgomery-Åsberg Depression Rating Scale
CMHT	community mental health team	MeSH	medical subject heading
CRHTT	Crisis Resolution and Home Treatment Team	MD	mean difference
Cri	credible interval	MDD	major depressive disorder
DIC	deviance information criterion	MTC	mixed-treatment comparison
DSM (-III, -IV)	<i>Diagnostic and Statistical Manual of Mental Disorders-Third and Fourth Edition</i>	NHS EED	NHS Economic Evaluation Database
ECG	electrocardiogram	NICE	National Institute for Health and Care Excellence
ECT	electroconvulsive therapy	NMS	neuroleptic malignant syndrome
EQ-5D	European Quality of Life-5 Dimensions	OD	once daily
ER	extended release	OR	odds ratio
		OWSA	one-way sensitivity analysis
		PICO	population, intervention, comparator and outcome

LIST OF ABBREVIATIONS

p.o.	per os	SG	standard gamble
PP	per protocol	SIP	Sickness Impact Profile
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses	SNRI	serotonin–norepinephrine reuptake inhibitor
PSA	probabilistic sensitivity analysis	SSRI	selective serotonin reuptake inhibitor
Q-LES-Q	Quality of Life Enjoyment and Satisfaction Questionnaire	STAR*D	Sequenced Treatment Alternatives to Relieve Depression
QALY	quality-adjusted life-year	TMS	transcranial magnetic stimulation
QoL	quality of life	TRD	treatment-resistant depression
QUOROM	Quality of Reporting of Meta-analyses	TSD	Technical Support Document
QWB	Quality of Well-Being Scale	TTO	time trade-off
QWB-SA	self-administered Quality of Well-Being Scale	WFSBP	World Federation of Societies of Biological Psychiatry
RCT	randomised controlled trial	WHOQOL-BREF	World Health Organization Quality of Life Assessment
RR	relative risk/risk ratio		
SD	standard deviation		
SE	standard error		
SF-36	Short Form questionnaire-36 items		

Scientific summary

Background

Patients with treatment-resistant depression (TRD) are those with major depressive disorder (MDD) that has not responded adequately to treatment. However, there is much uncertainty regarding what constitutes the definition of TRD and whether or not, for example, a patient with a failure to respond to two antidepressants from the same class could be defined as treatment resistant. The focus of this review is patients with unipolar TRD and, for the purposes of this report, TRD has been defined as a failure to respond to two or more antidepressants in the current episode of depression.

No UK-specific data on the incidence or prevalence of TRD are available in the literature. However, it is understood that up to two-thirds of patients diagnosed with MDD will have a suboptimal response to first-line treatment with antidepressant drugs. The World Federation of Societies of Biological Psychiatry guidelines for biological treatment of unipolar depressive disorders state that 'as many as 50% of non-responders to a first antidepressant trial also fail to respond to a second, different course of treatment'.

There are several strategies available for the treatment of patients with TRD. These strategies include pharmacological, non-pharmacological, and psychological and psychosocial interventions. Pharmacological treatment options include switching to a different antidepressant, the addition of another antidepressant of a different class, or use of an augmenting agent, such as anticonvulsant drugs, lithium or atypical antipsychotic drugs (AAPs).

This report contains a health technology assessment of lithium and AAPs used as augmentation therapies in the management of patients with TRD who are already taking selective serotonin reuptake inhibitor (SSRI) antidepressant therapy.

Objectives

The objective of the project was to estimate the clinical effectiveness and cost-effectiveness of augmentation of SSRI antidepressant therapy with either lithium or an AAP in the management of people with unipolar TRD.

The project was split into four distinct pieces of work:

- systematic review of clinical effectiveness of interventions
- systematic review of cost-effectiveness of interventions
- systematic review of quality-of-life (QoL) studies in depression
- de novo economic model.

Methods

Search methods

A systematic review of the literature was carried out to identify potentially relevant randomised controlled trials (RCTs) comparing augmentation of SSRI antidepressant therapy with either lithium or an AAP in the management of people with unipolar TRD. Databases searched were EMBASE, MEDLINE, PsycINFO and the Cochrane Central Register of Controlled Trials (CENTRAL). In addition, the registries of the Cochrane Collaboration Depression, Anxiety and Neurosis Review Group were searched. The NHS Economic Evaluation

Database (NHS EED) was also searched for the economic evaluation studies. All searches were performed from the date of database inception to August 2011. Further data were obtained from manufacturers. Inclusion decisions, quality assessment and data extraction were undertaken independently by two reviewers according to predefined criteria.

Clinical effectiveness

Standard pairwise meta-analysis was conducted using a fixed-effects model as the primary analysis. Mixed-treatment comparisons (MTCs) were conducted using a fixed- and random-effects model, with the best fitting most appropriate model chosen for the reporting of results. The systematic review was registered on PROSPERO (CRD42011001464).

Cost-effectiveness

A de novo mathematical model was developed to synthesise the available data on costs and clinical outcomes from the UK NHS perspective. The model adopted a 1-year time horizon, consisting of 8 weeks of acute treatment (captured by a decision tree) and 10 months of maintenance treatment (captured by a Markov model). The primary outcome of interest was the cost per quality-adjusted life-year (QALY) gained. The model required data on the outcomes of remission, response and discontinuation. Acute efficacy data used in the model were derived from the MTC carried out as part of this review. As a result of a paucity of clinical effectiveness data, a novel sampling approach was used to generate the probabilities required for the economic model. The approach involved sampling the treatment effect [change in MADRS (Montgomery–Åsberg Depression Rating Scale) score from baseline] of each augmentation strategy (from a distribution of possible effects) and calculating the proportion of patients (in a cohort of 1000 for each treatment arm) that would achieve remission or response during the acute treatment phase. The model assumed that outcomes in the maintenance phase were treatment independent.

Results

Clinical effectiveness

Twelve RCTs were identified in the review of clinical effectiveness data. Ten RCTs considered SSRI + AAP compared with SSRI + placebo/no treatment. Of the remaining two RCTs, one was a comparison of SSRI or serotonin–norepinephrine reuptake inhibitor (SNRI) + AAP with SSRI or SNRI + lithium and the final RCT compared SSRI + lithium with SSRI + placebo. Six of the 10 SSRI + AAP trials were included in the primary analysis; the remaining four RCTs were included in a class-based sensitivity analysis. Of the trials considering lithium augmentation as a comparator, only one was included in the primary analysis. All six trials considering augmentation with an AAP included in the primary analysis evaluated fluoxetine (SSRI) + olanzapine (AAP). Furthermore, the lithium trial included in the primary analysis used fluoxetine as the background SSRI in both the comparator group and lithium augmentation group.

Results for selective serotonin reuptake inhibitor plus atypical antipsychotic compared with selective serotonin reuptake inhibitor alone (pairwise comparison)

Five RCTs reported response based on the MADRS and the remaining RCT used the Hamilton Depression Rating Scale. The results of the meta-analysis (fixed effects) demonstrated a statistically significant benefit for fluoxetine + olanzapine over fluoxetine alone [odds ratio (OR) 1.48; 95% confidence interval (CI) 1.13 to 1.94] with a moderate level of statistical heterogeneity ($I^2 = 53\%$; $p = 0.07$).

Five RCTs reported the outcome of remission. Fixed-effects meta-analysis demonstrated a statistically significant increase in remissions in patients treated with olanzapine + fluoxetine compared with fluoxetine alone (OR 1.77; 95% CI 1.27 to 2.47) with no statistical heterogeneity ($I^2 = 0\%$; $p = 0.75$). Data on relapse rates were not available for analysis.

Four RCTs reported least square mean difference (MD) from baseline in MADRS score at study end point. Fixed-effects meta-analysis resulted in a statistically significant MD of -2.04 (95% CI -3.25 to -0.82) in favour of fluoxetine + olanzapine. However, there was a high level of heterogeneity that was statistically significant ($I^2 = 73\%$; $p = 0.01$).

Fixed-effects meta-analysis of five trials found that olanzapine augmentation therapy was associated with a non-statistically significant increase in discontinuations (OR 1.25; 95% CI 0.91 to 1.71) with no statistical heterogeneity ($I^2 = 0\%$; $p = 0.51$).

Selective serotonin reuptake inhibitor plus lithium compared with selective serotonin reuptake inhibitor plus placebo response (pairwise comparison)

The single trial comparing fluoxetine + lithium with fluoxetine alone used two definitions of response, one prespecified primary analysis and one post hoc analysis. Results of the primary and post hoc analyses for response data indicated a non-significant trend in favour of lithium augmentation compared with SSRI alone (OR 1.48; 95% CI 0.37 to 5.95 and OR 3.85; 95% CI 0.80 to 18.62, respectively). Data on remission or relapse rates were not available.

The MD in change in MADRS score from baseline between fluoxetine + lithium compared with fluoxetine alone was -3.79 (95% CI -11.25 to 3.67) – a non-significant improvement from baseline score with fluoxetine + lithium compared with fluoxetine alone.

Data on all-cause withdrawals demonstrated fewer withdrawals with a lithium augmentation strategy than with fluoxetine alone, although this difference was statistically non-significant (OR 0.68; 95% CI 0.15 to 3.16).

Mixed-treatment comparison (selective serotonin reuptake inhibitor plus atypical antipsychotic compared with selective serotonin reuptake inhibitor plus lithium)

Seven RCTs were included in the MTC: six for SSRI + AAP compared with SSRI alone and one RCT for SSRI + lithium compared with SSRI alone. Two separate analyses for the outcome of response were conducted because the trial informing the comparison with lithium reported response using two criteria. Analyses of response (random-effects model) using the lithium primary analysis and the lithium post hoc analysis data showed a non-significant trend in favour of treatment with lithium [OR 1.29; 95% credible interval (CrI) 0.11 to 5.32 and OR 4.15; 95% CrI 0.25 to 20.34, respectively].

Five trials were included in the analysis for mean change in MADRS (four RCTs were AAPs and one was a lithium RCT). The random-effects model resulted in a weighted MD of -1.47 (95% CrI -9.10 to 6.41) for the mean change in MADRS score from baseline for fluoxetine + lithium compared with fluoxetine + olanzapine, which suggests a statistically non-significant trend in favour of lithium augmentation. However, the wide 95% CrI indicates a high level of uncertainty in this estimate of treatment effect and so the results should be interpreted with caution.

Six trials reported data on all-cause withdrawals. The fixed-effects model results suggested a statistically non-significant trend in favour of augmentation with lithium (OR 0.74; 95% CrI 0.10 to 2.66) compared with augmenting with AAP.

Various sensitivity analyses were carried out, including analyses assuming class effects of SSRIs and AAPs, analysis of RCTs in which patients had experienced two or more failures to antidepressants in their current episode, and analysis of RCTs reporting response based on MADRS score. Results of most sensitivity analysis were consistent with the results of the primary analysis. However, the result of the sensitivity analysis assuming a class effect for SSRIs and AAPs for the outcome of mean change in MADRS differed from the primary analysis, identifying a statistically non-significant trend in favour of treatment with SSRI + AAP [MD 1.27; 95% CrI -1.88 to 4.68 (random-effects model)].

Cost-effectiveness

The systematic literature review identified four economic evaluations in the management of TRD and five studies that reported utility values for different levels of depression severity and treatment response. Of the economic evaluations, none directly addressed the review question but all were used to inform the modelling methods. Of the health-state utility values studies, one was used in the QALY calculations of the de novo model.

The monthly cost of the commonly prescribed SSRIs and AAPs varied substantially, whereas the costs of monitoring were modest. The annual cost per patient treated with SSRI + lithium was estimated to be £4739 compared with £5644 for those treated with SSRI + AAP. The difference in cost between the two augmentation strategies is around £905 per person per year, in favour of augmentation with lithium (lithium augmentation is cheaper) and translates to savings of £75 per person per month.

The results of the de novo modelling indicate that augmentation of SSRI therapy with an AAP is dominated by augmentation of an SSRI with lithium. The difference in costs is modest (cost savings £905 per person per year) and the difference in QALYs is estimated to be 0.03 QALYs. It appears there is no uncertainty about the dominance result, as lithium augmentation provided more benefits than AAP augmentation in all probabilistic runs. One-way sensitivity analysis showed that changes in costs had a minimal impact on the overall results, whereas changes in acute efficacy or discontinuation could potentially reverse the direction of the cost-effectiveness results. For example, assuming a low level of acute response (i.e. using the upper CrI) for the treatment effect of lithium (vs. SSRI alone; MD – 12.58, 95% CrI – 33.0 to 7.84) resulted in AAP augmentation having incremental cost-effectiveness ratios of < £20,000/QALY compared with lithium augmentation, as opposed to the base case in which lithium dominates. A similar result was observed when a high level of acute response for AAP was assumed (i.e. using the lower CrI) for AAP treatment effect (vs. SSRI alone; MD – 11.22, 95% CrI – 30.13 to 7.69). However, lithium remained dominant in most of the sensitivity analyses performed.

Discussion

The major weakness of this analysis is the lack of head-to-head data on the effectiveness of the comparison of SSRI + AAP with SSRI + lithium in patients with TRD. The MTC results demonstrate a non-significant benefit of augmenting with lithium compared with augmenting with AAP for most of the outcomes assessed. These results should be interpreted with caution, as the definitions of response and characteristics of patients included in the lithium RCTs differed from those used in the AAP RCTs. The inclusion of slightly different populations and definition of response may bias the results in favour of lithium; however, it should be noted that extent of the bias remains unknown.

The economic analysis found that augmenting with lithium was cheaper and more effective than augmenting with AAP. However, these results should be interpreted with caution in light of the sensitivity of the model to changes in the efficacy parameters and the uncertainty around the clinical data. In addition, it is expected that the price of AAPs will fall once generic versions of branded treatments are available. It is also important to note that the trials included in the MTC did not report on all of the outcomes of interest for the economic model and consequently some of the required parameters were generated using sampling methods that have not been previously validated. There was also a paucity of RCT data on follow-up and maintenance treatment.

Conclusions

The results of this review support the conclusion that augmentation of SSRIs with lithium or AAP is likely to be beneficial in people with TRD, defined as a failure to respond to two or more antidepressants in the current episode of depression. However, based on the limited number of RCTs identified, the clinical

evaluation suggests there is no statistically significant difference between the two augmentation strategies. There is a general paucity of trial data available in patients with TRD for SSRI + lithium and SSRI + AAP.

The cost-effectiveness results suggest that augmentation with lithium is cheaper and more effective compared with augmenting with AAP. However, the results are not definitive because the model is sensitive to the clinical effectiveness parameters of discontinuations and treatment response. The cost-effectiveness of SSRI + lithium and SSRI + AAP will need to be reconsidered if further trial data become available.

Suggested research priorities

A RCT in patients with TRD that compares SSRI + lithium with SSRI + AAP for response, remission and discontinuation in both the acute and maintenance phases of treatment is needed. In addition, data on relapse rates in the long term would be beneficial. Adverse events and QoL data should also be prioritised as part of the research.

Study registration

This study is registered as PROSPERO CRD42011001464.

Funding

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

Chapter 1 Background

Description of health problem

The focus of this review is the acute-phase treatment of patients with unipolar treatment-resistant depression (TRD). Patients with TRD are those with major depressive disorder (MDD) that has not responded adequately to treatment. However, there is much uncertainty regarding what constitutes the definition of TRD and whether or not, for example, a patient with a failure to respond to two antidepressants from the same class could be defined as treatment resistant.^{1,2} For the purposes of this report, TRD has been defined as a failure to respond to two or more antidepressants in the current episode of depression, the definition used in the 2003 National Institute for Health and Care Excellence (NICE) clinical guideline on the treatment and management of depression in adults (CG23).³ This definition for TRD was also reported by the authors of a large systematic review of 42 randomised trials to reflect the consensus within the literature (26 trials)⁴ and in the World Federation of Societies of Biological Psychiatry (WFSBP) guidelines⁵ for biological treatment of unipolar depressive disorders. In addition, the WFSBP guidelines⁵ state that 'as many as 50% of non-responders to a first antidepressant trial also fail to respond to a second, different course of treatment'. However, there is a general lack of clarity or consensus around the length of treatment required prior to treatment being defined as a failure and also the impact of historical treatment failures on the definition of future episodes of TRD, i.e. whether or not TRD should be diagnosed based on antidepressant failures that have occurred in only the current episode of depression.

Pathophysiology

The aetiology of depression is not fully understood, although there is evidence to suggest depression is a complex interaction among biological, genetic, psychosocial and environmental factors.² The highest rates of depression typically occur in people between 25 and 44 years old, and females are twice as likely as males to experience depression,² although how these figures relate to the subgroup of patients with TRD is difficult to know owing to the lack of epidemiological data and the lack of a consistent definition for TRD. Family history of depression is also a risk factor for depression² and a previous history of MDD increases the risk of future episodes (i.e. relapses).⁶ In addition, it has been reported that patients with depression have increased morbidity and mortality.² For example, they are more likely to die from cardiovascular disease⁷ or suicide.⁸

Diagnosis and assessment of response to treatment

People presenting with depression may complain of depressed mood, loss of interest or pleasure, feelings of guilt or low self-worth, suicidal ideation, disturbed sleep or appetite, low energy and poor concentration. Depression can be diagnosed clinically using different criteria. The most commonly used criteria are the DSM-IV (*Diagnostic and Statistical Manual of Mental Disorders-Fourth Edition* published by the American Psychiatric Association)⁹ and the ICD-10 (*International Classification of Diseases, Tenth Revision*) criteria developed by the World Health Organization (WHO).¹⁰ The DSM-IV system requires at least five out of nine symptoms for a diagnosis of major depression, including at least one of the following two symptoms: low mood; or loss of interest and pleasure.⁹ A diagnosis of moderately severe depressive episode using the ICD-10 classification system requires the presence of at least three out of ten depressive symptoms, including at least two of the following three symptoms: low mood; loss of interest and pleasure; or loss of energy.¹⁰ In both cases, symptoms should be present for at least 2 weeks and each symptom should be present at sufficient severity for most of every day (*Box 1* shows the full diagnostic criteria).

The recommended treatment goal in depression is to reach remission, which is defined as the relative absence of clinical symptomatology and is usually determined by reaching a certain score on a treatment response rating scale.^{11,12} Response to treatment in TRD is commonly measured by a reduction of at least 50% on either the Hamilton Depression Rating Scale (HAM-D)¹³ or Montgomery–Åsberg Depression

BOX 1 *Diagnostic and Statistical Manual of Mental Disorders-Fourth Edition*⁹ and ICD-10¹⁰ diagnostic criteria**The DSM-IV diagnostic criteria for major depressive disorder**

- Depressed mood most of the day, nearly every day, as indicated either by subjective report (e.g. feels sad or empty) or observation made by others (e.g. appears tearful).
- Markedly diminished interest or pleasure in all, or almost all, activities most of the day, nearly every day (as indicated either by subjective account or observation made by others).
- Significant weight loss when not dieting or weight gain (e.g. a change of > 5% of body weight in a month), or decrease or increase in appetite nearly every day.
- Insomnia or hypersomnia nearly every day.
- Psychomotor agitation or retardation nearly every day (observable by others, not merely subjective feelings of restlessness or being slowed down).
- Fatigue or loss of energy nearly every day.
- Feelings of worthlessness or excessive or inappropriate guilt (which may be delusional) nearly every day (not merely self-reproach or guilt about being sick).
- Diminished ability to think or concentrate, or indecisiveness, nearly every day (either by subjective account or as observed by others).
- Recurrent thoughts of death (not just fear of dying), recurrent suicidal ideation without a specific plan, or a suicide attempt or specific plan for committing suicide.

The ICD-10 diagnostic criteria for moderate depressive episode

- persistent sadness or low mood; *and/or*
- loss of interests or pleasure
- fatigue or low energy
- disturbed sleep
- poor concentration or indecisiveness
- low self-confidence
- poor appetite
- suicidal thoughts or acts
- bleak and pessimistic views of the future
- guilt or self-blame.

Rating Scale (MADRS).¹⁴ Both scales are designed to be administered via a clinical interview and consist of a list of symptoms of depression which the clinician must assess the patient for during the interview. The clinician then rates the patient's symptoms on the scales provided for each symptom and adds up the individual scores to provide the overall score. In both cases, the lower the score, the less severely depressed the patient. The definition of remission on the HAMD is typically defined as a score of ≤ 7 on the 17-item version of the HAMD.¹⁵ However, there is currently no consensus on a definition of remission for the MADRS,¹⁵ although clinical expert advisors reported that scores of ≤ 10 on the MADRS are commonly used.

Incidence and prevalence

The current NICE clinical guideline on the treatment and management of depression in adults (CG90)¹⁶ states that there are 130 people with depression per 1000 of the NHS population, although only 80 people per 1000 of the population actually consult their general practitioner (GP). A survey carried out by the Social Survey Division of the Office for National Statistics on behalf of the Department of Health, the Scottish Executive and the National Assembly for Wales in 2000¹⁷ reported a prevalence of the diagnosis of depressive

disorder of 28 per 1000 of the survey population. A comparison in the report¹⁷ was made to an earlier survey conducted in 1993, in which the prevalence of depressive disorder was reported as 23 per 1000 people, thus suggesting that either the prevalence of depression in the UK is rising or that depression is being diagnosed more frequently. A study looking at depression in England in the adult population in 2000¹⁸ reported the total number of people suffering from depression in England to be 2,661,468, based on calculations using 1998 data from the Office for National Statistics. The study also reported that 72% of people with depression were female, and 20% were in the 35- to 44-year age band.¹⁸ In CG90¹⁶ it is reported that the prevalence of depression also varies considerably according to sex and a wide range of social and economic factors (e.g. it is higher in females and in unemployed people). However, it should be noted that the statistics reported above reflect the total clinical spectrum of depression and, as such, encompass patients with mild, moderate and severe depression, and thus those with TRD represent only a subgroup of these. No data specific to the UK incidence or prevalence of TRD were identified, which is possibly because there is no widely agreed definition for TRD.¹

Current service provision

There are several pharmacological treatment strategy options for patients with TRD not achieving adequate response with antidepressants. In the *British National Formulary* (BNF)¹⁹ it is stated that:

Failure to respond to initial treatment with an SSRI [selective serotonin reuptake inhibitor] may require an increase in the dose, or switching to a different antidepressant. Failure to respond to a second antidepressant may require the addition of another antidepressant of a different class, or use of an augmenting agent [such as lithium, aripiprazole (unlicensed), olanzapine (unlicensed), quetiapine or risperidone (unlicensed)], but such adjunctive treatment should be initiated only by doctors with special experience of these combinations.

Other potential treatment options suggested in CG90¹⁶ include augmenting with a different agent, such as anticonvulsants, pindolol (Visken®, Amdipharm), triiodothyronine, benzodiazepines, buspirone or atomoxetine (Strattera®, Eli Lilly), or the use of electroconvulsive therapy (ECT) or psychological and psychosocial interventions such as cognitive-behavioural therapy.

Relevant national guidelines

The key clinical guideline for depression in adults in the UK is the NICE clinical guideline on the treatment and management of depression in adults (CG90; extracts from this guideline have been reproduced here with permission);¹⁶ this guideline is the updated version of CG23.³ It should be noted that in CG90¹⁶ a decision was taken to no longer use the term 'treatment-resistant depression', as there were concerns that the term implies there is a 'natural cut-off between people who respond to one or two antidepressants compared with those who do not', and that this 'is not supported by the evidence, and the term may be taken by both doctors and patients as a pejorative label'.¹⁶ As a result, in CG90¹⁶ it was decided to approach the problem of inadequate response by considering sequenced treatment options rather than by a category of patient. This is reflected throughout CG90¹⁶ by use of the label 'inadequate response to initial interventions'.

The recommendations in CG90¹⁶ for the sequencing of drug treatments in patients with an initial inadequate response are presented in *Box 2*, together with the additional recommendations for monitoring when treatment with lithium or atypical antipsychotic drugs (AAPs) is chosen.

In addition to the NICE clinical guideline (CG90) for the treatment and management of depression in adults,¹⁶ there are guidelines published by the British Association of Psychopharmacology (BAP).²⁰ The BAP 2008 guidelines²⁰ provide similar advice to CG90¹⁶ on the use of augmentation therapy in TRD, although they do not specifically mention a definition of how many treatment failures are required for a diagnosis of TRD (*Box 3*).

BOX 2 The NICE guideline, CG90.¹⁶ National Institute for Health and Care Excellence (NICE). *Depression: the treatment and management of depression in adults (update)*. CG90. London: NICE; 2009. URL: <http://guidance.nice.org.uk/CG90>. Reproduced with permission

1.8.1.1 When reviewing drug treatment for a person with depression whose symptoms have not adequately responded to initial pharmacological interventions:

- check adherence to, and side effects from, initial treatment
- increase the frequency of appointments using outcome monitoring with a validated outcome measure
- be aware that using a single antidepressant rather than combination medication or augmentation is usually associated with a lower side effect burden
- consider introducing previous treatments that have been inadequately delivered or adhered to, including increasing the dose
- consider switching to an alternative antidepressant.

Combining and augmenting medications

'Augmentation' is when an antidepressant is used with a non-antidepressant drug and 'combination' is when two antidepressants are used together.

1.8.1.6 If a person with depression is informed about, and prepared to tolerate, the increased side effect burden, consider combining or augmenting an antidepressant with:

- lithium *or*
- an antipsychotic such as aripiprazole,^a olanzapine,^a quetiapine^a or risperidone,^a *or*
- another antidepressant such as mirtazapine or mianserin.

1.8.1.7 When prescribing lithium:

- monitor renal and thyroid function before treatment and every 6 months during treatment (more often if there is evidence of renal impairment)
- consider ECG [electrocardiogram] monitoring in people with depression who are at high risk of cardiovascular disease
- monitor serum lithium levels 1 week after initiation and each dose change until stable, and every 3 months thereafter.

^a In this guideline, drug names are marked with an asterisk if they do not have UK marketing authorisation for the indication in question at the time of publication (October 2009).

A key problem highlighted in both of the guidelines^{16,20} is that there is limited randomised controlled trial (RCT) evidence comparing the different potential augmentation treatments and thus there is currently much uncertainty as to which augmentation therapy is the most clinically effective and/or cost-effective in the management of TRD.

Current service cost

No data were identified that reported specifically on the economic impact of TRD in the UK. However, the report of a survey carried out by the Social Survey Division of the Office for National Statistics in 2000¹⁷ estimated the total cost of adult depression to be over £9B, including around £370M of direct treatment costs. It also estimated that there were 109.7 million working days lost and 2615 deaths as a result of

BOX 3 The BAP 2008 guidelines²⁰**Augmentation/combination treatment (A)**

- Consider adding a second agent especially if:
 - there is partial/insufficient response on the current antidepressant (D) and
 - there is good tolerability of current antidepressant (D);
 - switching antidepressant has been unsuccessful (D)
- establish the safety of the proposed combination (S)
- choose the combinations with the best evidence-base first (S)
- consider adding lithium (A), olanzapine (A), quetiapine (B), risperidone (B), aripiprazole (B), triiodothyronine (B) or mirtazapine (B) being aware that the evidence mainly supports lithium and triiodothyronine added to TCAs [tricyclic antidepressants] and the other drugs added to SSRIs.

Notes:

1. Developed from Shekelle *et al.* 1999.

2. Categories of evidence for causal relationships and treatment:

- Evidence from meta-analysis of randomised controlled trials,^a at least one large, good quality, randomised controlled trial^a or replicated, smaller, randomised controlled trials*
- Evidence from small, non-replicated, randomised controlled trials,^a at least one controlled study without randomisation or evidence from at least one other type of quasi-experimental study
- Evidence from non-experimental descriptive studies, such as uncontrolled, comparative, correlation and case-control studies
- Evidence from expert committee reports or opinions and/or clinical experience of respected authorities

3. Proposed categories of evidence for non-causal relationships:

- Evidence from large representative population samples
- Evidence from small, well designed, but not necessarily representative samples
- Evidence from non-representative surveys, case reports
- Evidence from expert committee reports or opinions and/or clinical experience of respected authorities

4. Strength of recommendation:

- A. Directly based on category I evidence
- B. Directly based on category II evidence or extrapolated^b recommendation from category I evidence
- C. Directly based on category III evidence or extrapolated^b recommendation from category I or II evidence
- D. Directly based on category IV evidence or extrapolated^b recommendation from category I, II or III evidence
- S. Standard of good practice.

a Randomised controlled trials must have an appropriate control treatment arm; for primary efficacy this should include a placebo condition.

b Extrapolation may be necessary when evidence is only indirectly related, covers only a part or the area of practice under consideration, or has methodological problems or is contradictory.

depression in 2000.¹⁷ These figures represent the whole spectrum of depression and so the actual costs of TRD are likely to be much lower, although it could reasonably be expected that the costs per patient would be significantly higher for patients with TRD compared with the costs of treatment of mild depression. This assumption is supported by reports that patients with TRD use a disproportionately larger share of health-care resources and cost employers more in lost productivity than patients with MDD who respond to treatment.^{2,21}

A more recent economic review was conducted by the King's Fund in 2006 to estimate expenditure on mental health, including depression, in England up to 2026.²² This study²² estimated the total costs for depression, including prescribed drugs, inpatient care, other NHS services, supported accommodation, social services and lost employment. The total cost of services for depression in England in 2007 was estimated to be £1.7B, with costs projected to reach £3B by 2026.²² Moreover, addition of costs attributable to lost employment increased the estimated cost of depression in 2007 to £7.5B (projected to reach £12.2B by 2026).²² These figures are consistent with a more recent report by the UK Mental Health Foundation in November 2010;²³ this reported that depression costs 'the UK economy over £9 bn a year in lost earnings – an increase of £4 bn since 1999, and a rise of over £500 m in the last year alone'. The figures for the UK Mental Health Foundation report²³ were calculated by the Research Service of the House of Commons Library. The estimates identified for the costs of depression are all consistent with a statement in CG90¹⁶ that the indirect costs of depression far outweigh the health service costs.

Description of technology under assessment

The technologies under assessment in this report are lithium and AAPs used as augmentation therapies in the management of patients with TRD taking concomitant SSRI antidepressant therapy. All of the treatments under investigation in this review are available as oral tablet or liquid formulations.¹⁹ Some of the AAPs (aripiprazole, olanzapine and risperidone) are also available for administration parenterally (e.g. intravenously or intramuscularly). The current usage in the NHS of lithium and AAPs as augmentation therapies in TRD varies across different regions. This is likely to be due to the absence of national guidelines or treatment pathways recommending a preference for either augmentation strategy for TRD. The NICE clinical guideline on depression in adults (CG90)¹⁶ recommends the following AAPs to augment the effectiveness of SSRIs in TRD: aripiprazole (Abilify®, Bristol-Myers Squibb); olanzapine (Zyprexa®, Eli Lilly); quetiapine (Seroquel®, AstraZeneca); and risperidone (Risperdal®, Janssen).

The augmentation therapies that are the focus of this report have been evaluated in patients who have failed to respond to two or more antidepressants in their current episode of depression. Clinical advisors for this report have suggested that augmentation with lithium or an AAP in TRD may be commenced after either one or two antidepressant failures in the current episode of depression, depending on the patient's medical history and current clinical status. However, the clinical advisors reported that, in their experience, as many as 50% of patients who fail on an initial SSRI will respond to a second SSRI. The experts thus consider that the population of patients with failure to respond to two or more previous antidepressants in their current episode of depression represents an appropriate population in which to evaluate the clinical effectiveness and cost-effectiveness of augmentation with either lithium or AAP.

The NICE clinical guideline on depression in adults (CG90)¹⁶ reports on the use of augmentation agents after an initial inadequate response to treatment of depression, although this is not further defined. In these patients, CG90¹⁶ recommends that augmentation with lithium or an AAP could be considered as a potential treatment strategy. In addition, it is considered by the clinical experts for this report that, at this time, lithium is used less frequently in the NHS than AAPs in the treatment of patients with TRD.

The duration of augmentation therapy is variable, and is partly dependent on the length of time until remission is reached. For this review, it is anticipated that treatment with any augmentation agent should be for a minimum of 4 weeks prior to the final efficacy assessment, and treatment with AAPs in particular is

recommended for a minimum of 4 weeks before discontinuation due to lack of efficacy.¹⁹ Usually, it is expected that augmentation therapy would be discontinued after a period of time in remission, although there is currently no set duration for this maintenance of treatment. Clinical advisors for this report suggest that treatment should be maintained for around 6 months post diagnosis of remission. CG90¹⁶ also suggests that treatment should be continued for at least 6 months after remission, and in patients at risk of relapse treatment should be continued for at least 2 years. In clinical practice, augmentation of antidepressants may occur in primary, secondary or tertiary care, and is usually in an outpatient setting. In CG90¹⁶ it is recommended that augmentation therapy should be started only in primary care in consultation with a consultant psychiatrist. Clinical experts for this report estimate that approximately 70% of patients with TRD will receive care from their GP and a community mental health team (CMHT). Of the remaining patients, it is estimated that 20% will be seen by Crisis Resolution and Home Treatment Teams (CRHTTs), which provide intensive home-based support, and the remainder of patients would receive inpatient care. Follow-up for patients with TRD is usually dependent on the patient's clinical need and also the requirement for monitoring associated with the individual augmentation therapy, which is discussed in more detail below.

Selective serotonin reuptake inhibitors

The SSRIs licensed for use in the UK are citalopram, escitalopram, fluoxetine, fluvoxamine, paroxetine and sertraline. They are all available in England and Wales as both generic and branded drugs: citalopram (Cipramil[®], Lundbeck); escitalopram (Cipraxel[®], Lundbeck); fluoxetine (Prozac[®], Lilly); fluvoxamine (Faverin[®], Abbott Healthcare); paroxetine (Seroxat[®], GlaxoSmithKline); and sertraline (Lustral[®], Pfizer). SSRIs work by selectively inhibiting the reuptake of serotonin [5-hydroxytryptamine (5-HT)]; hence, they are termed SSRIs. SSRIs are commonly used first-line for treating depression as they are better tolerated and are safer in overdose than other classes of antidepressants.¹⁹ In particular, the SSRIs are less sedating and have fewer antimuscarinic and cardiotoxic effects than tricyclic antidepressants.¹⁹ Side effects of SSRIs include gastrointestinal effects (e.g. nausea, vomiting, abdominal pain, diarrhoea, constipation), anorexia, rash, dry mouth, anxiety, headache, insomnia, tremor, dizziness, asthenia, drowsiness, convulsions, sexual dysfunction, urinary retention, sweating and hyponatraemia.¹⁹

SSRIs are recommended for use with caution in patients with epilepsy, cardiac disease, diabetes mellitus, susceptibility to angle-closure glaucoma, a history of mania or bleeding disorders, or taking drugs that increase the risk of bleeding.¹⁹ SSRIs are associated with a risk of seizures and should therefore be used with caution in those receiving concurrent ECT.¹⁹

Lithium

Lithium is used in the UK to augment antidepressants in patients with TRD, although this is an unlicensed indication.¹⁹ Lithium is more commonly used for its licensed indication as a mood-stabilising agent, although the precise mechanism of action of lithium remains unknown.

Lithium is available in both generic and branded formulations in England and Wales. These are as follows: lithium carbonate (Camcolit[®], Norgine; Lithonate[®], Teva UK; Liskonum[®], GlaxoSmithKline; Priadel[®] tablets, Sanofi-aventis); and lithium citrate (Li-Liquid[®], Rosemont; Priadel[®] liquid, Sanofi-aventis).

Lithium salts have a narrow therapeutic–toxic ratio and therefore serum lithium concentrations should be monitored regularly during treatment.¹⁹ Levels should be measured 12 hours after the dose with the aim of achieving a serum-lithium concentration of 0.4–1 mmol/l.¹⁹ The BNF recommends that routine 'serum-lithium monitoring should be performed weekly after initiation and after each dose change until concentrations are stable, then every 3 months thereafter'.¹⁹ It should also be noted that different lithium preparations have different bioavailability and so caution is required when changing the lithium preparation.¹⁹

Renal function should also be monitored at baseline and every 6 months thereafter as lithium is excreted renally, and so renal impairment could cause lithium levels to build up, leading to toxicity.¹⁹ Serum lithium levels are also affected by a patient's sodium or fluid intake, with the risk of lithium toxicity increasing if

there is sodium depletion or dehydration. Long-term use of lithium has been associated with thyroid disorders and mild cognitive and memory impairment.¹⁹ Therefore, long-term treatment requires careful assessment of risk and benefit, and monitoring of thyroid function every 6 months.¹⁹

Side effects of lithium therapy include gastrointestinal disturbances, fine tremor, renal impairment, polydipsia, leucocytosis, weight gain, cognitive dulling, hyperparathyroidism, hyperthyroidism, hyperglycaemia, hypermagnesaemia and hypercalcaemia.¹⁹ The signs of lithium toxicity include blurred vision, anorexia, vomiting, diarrhoea, muscle weakness, polyuria and increasing drowsiness eventually leading to coma.¹⁹

It is recommended that lithium is avoided if possible in patients with renal impairment and used with caution in patients with cardiac disease; QT-interval prolongation; conditions with sodium imbalance (e.g. Addison's disease); diarrhoea; vomiting; intercurrent infection; concurrent ECT treatment; psoriasis; and myasthenia gravis.¹⁹ In addition, caution is recommended in the use of lithium in surgical patients, patients on diuretics and the elderly.¹⁹

Atypical antipsychotic drugs

As discussed above for lithium, AAPs can similarly be used as adjunctive therapies to antidepressants in the treatment of patients with TRD. AAP drugs are also known as the second-generation antipsychotic drugs and act on a range of receptors in comparison with the first-generation antipsychotic drugs that predominantly act on only one type of receptor. The only AAP licensed for use in the UK as an adjunctive treatment in MDD is quetiapine.¹⁹ However, there are several other AAPs that are used as unlicensed treatments in patients with unipolar TRD, including aripiprazole, olanzapine and risperidone. In addition, there are other AAPs classed as second-generation antipsychotic drugs that could also potentially be used, albeit unlicensed.

Most of the AAPs are still patented and thus are available only as branded drugs. The AAP drugs available for use in England and Wales are as follows: amisulpride (Solian[®], Sanofi-aventis); aripiprazole (Abilify[®], Bristol-Myers Squibb); clozapine (Clozaril[®], Novartis; Denzapine[®], Merz; Zaponex[®], Teva UK); olanzapine (Zyprexa[®], Lilly); paliperidone (Invega[®], AstraZeneca); quetiapine [Seroquel[®], AstraZeneca; Seroquel XL[®] (modified release), AstraZeneca]; risperidone (Risperdal[®], Janssen-Cilag). In addition to these AAPs, ziprasidone (Geodon[®]/Zeldox[®], Pfizer) is used elsewhere in Europe, but is not licensed or used routinely for any indication in the UK. As this report is focused on treatments available for use in the NHS, ziprasidone will not be discussed further in this section.

The choice of AAP medication is usually influenced by the patient's medication history, and consideration of individual patient factors, for example, the risk of particular side effects such as weight gain or impaired glucose tolerance. As previously discussed, AAPs each act on different receptors. These are summarised in *Table 1*.

TABLE 1 Mode of action of the AAPs¹⁹

AAP	Mechanism of action
Amisulpride	Selective dopamine receptor antagonist with high affinity for mesolimbic D ₂ and D ₃ receptors
Aripiprazole	Dopamine D ₂ partial agonist with weak 5-HT _{1a} partial agonism and 5-HT _{2A} receptor antagonism
Clozapine	Dopamine D ₁ , dopamine D ₂ , 5-HT _{2A} , alpha-1 adrenoceptor, and muscarinic-receptor antagonist
Olanzapine	Dopamine D ₁ , D ₂ , D ₄ , 5-HT ₂ , H ₁ - and muscarinic-receptor antagonist
Paliperidone	Metabolite of risperidone; dopamine D ₂ , 5-HT _{2A} , alpha-1 adrenoceptor, and H ₁ -receptor antagonist
Quetiapine	Dopamine D ₁ , dopamine D ₂ , 5-HT ₂ , alpha-1 adrenoceptor, and H ₁ -receptor antagonist
Risperidone	Dopamine D ₂ , 5-HT _{2A} , alpha-1 adrenoceptor, and H ₁ -receptor antagonist

Full blood count, urea and electrolytes, and liver function test monitoring are required at the start of therapy with antipsychotic drugs, and then annually thereafter.¹⁹ In addition, clozapine requires differential white blood cell monitoring weekly for 18 weeks then fortnightly for up to 1 year, and then monthly as part of the clozapine patient monitoring service.¹⁹ Blood lipids and weight should also be measured at baseline, at 3 months, and then yearly, and fasting blood glucose should be measured at baseline, at 4–6 months, and then yearly.¹⁹ Patients taking clozapine or olanzapine should have fasting blood glucose tested at baseline, after 1 month's treatment, and then every 4–6 months.¹⁹ It is also advisable to monitor prolactin concentration regularly. Blood pressure monitoring is also advised before starting therapy and frequently during dose titration of antipsychotic drugs, and ECG monitoring may also be required if the patient has cardiovascular risk factors.¹⁹

There are numerous side effects associated with AAP drugs and the side effects contribute significantly to the reasons for non-adherence to therapy.¹⁹

Most antipsychotic drugs increase prolactin concentration because dopamine inhibits prolactin release, but aripiprazole reduces prolactin because it is a dopamine-receptor partial agonist.¹⁹ Risperidone and amisulpride are most likely to cause symptomatic hyperprolactinaemia.¹⁹ The clinical symptoms of hyperprolactinaemia include sexual dysfunction, reduced bone mineral density, menstrual disturbances, breast enlargement and galactorrhoea.¹⁹

Other side effects associated with AAPs include cardiovascular side effects, such as tachycardia, arrhythmias and hypotension.¹⁹ Hyperglycaemia and sometimes diabetes can occur, particularly with clozapine, olanzapine, quetiapine and risperidone.¹⁹ All antipsychotic drugs may cause weight gain, although the risk and extent varies, with clozapine and olanzapine being the most commonly associated with weight gain.¹⁹ Clozapine and quetiapine can cause postural hypotension (especially during initial dose titration), which may be associated with syncope or reflex tachycardia in some patients.¹⁹ Hypersalivation is also associated with clozapine therapy.¹⁹ In addition, other possible side effects include drowsiness, agitation, restlessness, increased appetite, insomnia, dizziness, headache, confusion, gastrointestinal disturbances, venous thromboembolism, and antimuscarinic symptoms (e.g. dry mouth, constipation, difficulty with micturition, blurred vision and also, very rarely, precipitation of angle-closure glaucoma).¹⁹ Neuroleptic malignant syndrome (NMS) is a rare, but potentially fatal, side effect of all antipsychotic drugs and requires discontinuation of the antipsychotic drug.¹⁹ NMS is characterised by hyperthermia, a fluctuating level of consciousness, muscle rigidity, pallor, irregular pulse, tachycardia, sweating and urinary incontinence.

It is recommended that AAPs are used with caution in patients with cardiovascular disease, a history of epilepsy or those on concomitant drugs that increase the QT interval (on an ECG).¹⁹ In addition, caution is required in the elderly owing to an increased risk of mortality associated with antipsychotic drugs and an increased risk of other serious side effects.¹⁹

Anticipated costs associated with intervention

The direct costs associated with the interventions under review (SSRIs, AAPs and lithium) are limited to the price of the individual tablet or liquid formulations, as there is no requirement for them to be administered in a specialised setting. In addition, there are some costs associated with the monitoring requirements of each therapy, although these costs vary between lithium and AAPs, as well as among the individual AAPs. The costs of the interventions along with the wider costs associated with each intervention are discussed in detail in the cost-effectiveness section of this report (see *Chapter 5, Drug costs*).

Chapter 2 Definition of the decision problem

This section states the key factors that will be addressed by this report and defines the scope of the assessment (decision problem) in terms of these key factors, in line with the definitions agreed in the published project protocol (see *Appendix 1*).²⁴ The protocol for this systematic review was registered on PROSPERO, which is an international prospective register of systematic reviews (CRD42011001464);²⁵ the protocol is also available in full on the NIHR Health Technology Assessment (HTA) programme website (www.hta.ac.uk/project/2599).²⁴

Decision problem

This report aims to address the question ‘What is the clinical effectiveness and cost-effectiveness of lithium or an AAP in the management of treatment-resistant unipolar depression in adults?’

The planned population, intervention, comparator and outcomes (PICO) for this report was as follows:

- Population:
 - adults with treatment-resistant unipolar depression defined as failure to respond to at least two previous antidepressants in the current episode of depression.

Restrictions were not imposed on the maximum number of previous antidepressant drugs allowed in order to avoid reducing the amount of data available for analysis, as it was noted a priori that there may be limited relevant SSRI RCT data available. However, this decision assumed that there was a consistent relative treatment effect independent of line of therapy (i.e. addition of an AAP or lithium had the same relative benefit whether given with third-line SSRI or fourth-line SSRI, etc.) and so a sensitivity analysis was prespecified to assess the impact of this assumption.

- Intervention:
 - an SSRI (defined as citalopram, escitalopram, fluoxetine, fluvoxamine, paroxetine or sertraline), *plus*
 - an AAP drug (also known as second-generation antipsychotic, and defined as amisulpride, aripiprazole, clozapine, olanzapine, paliperidone, quetiapine, risperidone or ziprasidone).
- Comparator:
 - an SSRI (defined as citalopram, escitalopram, fluoxetine, fluvoxamine, paroxetine or sertraline), *plus*
 - lithium (lithium carbonate or lithium citrate or lithium).
- Outcomes:
 - response (measured by a reduction of at least 50% in HAMD¹³ or MADRS¹⁴ score)
 - remission (using individual trial definitions)
 - mean change from baseline MADRS¹⁴ score
 - quality of life (QoL) as reported using a validated QoL rating scale²⁵ [e.g. Short Form questionnaire-36 items (SF-36)]
 - adverse events (total number of events, and the individual adverse events deemed most burdensome to patients)
 - withdrawals (all cause) as a surrogate outcome for adherence to medication
 - relapse rate
 - mortality
 - cost-effectiveness.

Subgroup analyses

The a priori subgroup analyses deemed to be most important were as follows:

- different durations of depression (i.e. time since first onset of current episode of depression)
- class of previous antidepressants (e.g. SSRI or tricyclic antidepressant)
- sex (i.e. male and female)
- age (i.e. those < 75 years and those ≥ 75 years)
- people with different severities of depression (i.e. based on trial entry HAMD score¹³).

The reason for selecting these subgroups is that they were highlighted by clinical experts to be the most clinically important subgroups. This is because:

- People who have had TRD for longer periods of time are likely to be more difficult to treat and, thus, could be less likely to respond to augmentation therapy.
- Previous class of antidepressant therapy may have an impact on the response to future treatments (i.e. if two SSRIs have been failed in the current episode rather than two different classes of antidepressants).
- It is unknown whether or not sex has an effect on response to treatment in TRD, but more females tend to be treated for depression than males and thus RCTs may have a higher female–male ratio.
- People of < 75 years of age are known to have different pathophysiologies for their depression and also to respond differently to antidepressants than people aged ≥ 75 years.
- People with more severe depression at baseline (i.e. higher HAMD¹³ score) require a greater improvement to enter remission and so could potentially be less likely to enter remission.

It is thus considered that these subgroups of patients could respond differently to augmentation therapy and so each subgroup will be analysed for the primary outcome in this review if sufficient data are identified to enable such comparison.

Overall aims and objectives of assessment

The aim of this report is to compare the augmentation of SSRI antidepressant therapy with either lithium or an AAP in the management of people with treatment-resistant unipolar depression.

The key areas that this report plans to address are:

- identifying and reviewing the existing evidence relating to the clinical efficacy of augmentation of SSRIs with lithium or an AAP
- reporting the cost-effectiveness of augmentation of SSRIs with lithium compared with that of augmentation of SSRIs with an AAP
- identifying what the potential areas for future research might be in the pharmaceutical management of TRD.

Chapter 3 Assessment of clinical effectiveness

Methods for reviewing effectiveness

A review of the evidence for clinical effectiveness was undertaken systematically following the general principles recommended in the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) statement [formerly the QUOROM (Quality of Reporting of Meta-analyses) statement]²⁶ and based on the PICO defined in *Chapter 2*.

Search strategy

The search strategy comprised the following elements:

1. searching of electronic bibliographic databases
2. contact with clinical experts in the field
3. review of the reference lists of retrieved papers
4. searching of the Cochrane Collaboration Depression, Anxiety and Neurosis Review Group (CCDAN) Controlled Trials Register (CTR) databases (CCDANCTR-Studies and CCDANCTR-References).

The following electronic databases were searched:

- (a) EMBASE (searched from 1974 to August 2011)
- (b) MEDLINE (searched from inception to August 2011)
- (c) PsycINFO (searched from inception to August 2011)
- (d) Cochrane Central Register of Controlled Trials (CENTRAL) (searched from inception to August 2011).

Full details of the search strategies used are provided in *Appendix 2*.

In addition, to assist the drawing up of final recommendations, the website ClinicalTrials.gov was searched to identify relevant ongoing clinical trials. Trials considered relevant were those that when completed may have an impact on the results of this review. Clinical experts in the relevant therapy areas were contacted for details of trials (published and unpublished) of which they may be aware.

The references from any relevant review papers or RCTs uncovered in the search were also examined for additional references potentially relevant to the review.

The CCDANCTR-Studies and CCDANCTR-References were searched using terms consistent with the search terms used in the other electronic bibliographic databases as a validation exercise of the searches. The searches on the CCDANCTR-Studies and CCDANCTR-References registers were conducted on 7 December 2011.

Abstract appraisal

Titles and abstracts of studies identified by the search process were assessed independently by two reviewers (VH and SB) for inclusion. For cases in which the reviewers were unable to reach a consensus on whether or not the full text should be obtained for further appraisal, the full text was obtained.

When potentially relevant data were available in only an abstract format then attempts were made to contact the corresponding author and drug manufacturer to obtain the full publication or additional information if possible.

The a priori inclusion criteria applied to the review:

- for the review of clinical effectiveness, only RCTs were included
- adults ≥ 18 years
- people with unipolar depression
- TRD defined as failure to respond to at least two previous antidepressants in the current episode of depression
- SSRI given as baseline treatment and patient randomised to either lithium or an AAP
- comparator treatment of SSRI plus either lithium, AAP, placebo or no treatment
- minimum duration of 4 weeks' treatment with study medication for the current episode of depression
- studies reporting on one or more of the following outcomes:
 - response
 - QoL
 - adverse events
 - adherence to medication or withdrawals (all cause)
 - relapse rate
 - mortality
 - cost-effectiveness.

The a priori exclusion criteria applied to the review:

- non-randomised studies
- narrative reviews, editorials, opinions
- studies performed in animals
- studies not focusing on the treatment of the acute phase of depression (i.e. those focusing solely on maintenance therapy)
- bipolar depression or bipolar disorder diagnosis prior to study entry
- underlying medical condition or another substantial comorbid psychiatric condition (e.g. psychosis)
- trials reporting only post-crossover results
- trials using non-SSRI antidepressants as the baseline treatment for augmentation with lithium or an AAP.

Study inclusion assessment

Two reviewers (VH and SB) independently assessed the full-text papers of the trials identified during the abstract assessment stage for inclusion and any differences in opinion were arbitrated by a third reviewer (SJE).

Data extraction strategy

A sample of five papers was fully independently data extracted by two reviewers (VH and SB) using a standardised data extraction form (for a copy of the data collection form, please see *Appendix 3*) and then validated by one reviewer (SB). Agreement between the two reviewers was high and so, owing to time constraints, the remaining papers were independently extracted by one reviewer (VH) and validated by a second reviewer (SB). Discrepancies in the data extracted by the two reviewers were resolved through discussion, with involvement of a third reviewer (SJE) if necessary.

Data from intention-to-treat (ITT) analyses were extracted and it was planned that per-protocol (PP) data would also be extracted for use in a sensitivity analysis, although PP data were not reported in any of the included papers. For the purpose of this review, ITT was defined as patients being analysed in the treatment group to which they were allocated at randomisation regardless of whether they received the wrong intervention, withdrew or were lost to follow-up. Should a trial not report ITT data then missing data were treated as treatment failures to allow the analysis to conform to an ITT analysis.

Study authors and drug manufacturers were contacted to supply any additional information not included in published sources (including relevant subgroup data and additional methodological data required for the quality assessment).

Quality assessment strategy

Outcomes from the studies that met the inclusion criteria were assessed using the updated risk of bias tool developed by the Cochrane Collaboration (March 2011).²⁷

These criteria assess the following areas:

1. random sequence generation
2. allocation concealment
3. blinding of participants and personnel
4. blinding of outcomes assessment
5. incomplete outcome data
6. selective reporting
7. 'other bias'.

Based on these criteria, an assessment for each outcome reported in the trial was allocated based on the identified risk of bias. The three bias assessment categories used were low risk, high risk and unclear risk. Only trials that were deemed to be at low or unclear risk of bias were included in the main analysis, with plans to include the trials rated as high risk in a sensitivity analysis; no trial was rated as being at high risk of bias.

Two reviewers (VH and SB) independently rated the trial outcomes for inclusion and any differences in opinion were arbitrated by a third reviewer (SJE). Outcomes reported by each RCT were considered appropriate for inclusion unless the trial demonstrated a high risk of bias across several of the seven risk-of-bias domains assessed for that outcome. No trial was excluded from any of the outcomes analysed based on the risk of bias assessments.

Methods of analysis/synthesis

Data have been tabulated and, where appropriate, meta-analysis undertaken to estimate a summary measure of effect on relevant outcomes based on ITT analyses. Standard pairwise meta-analysis was conducted when more than one trial was identified for inclusion for any pair of treatments under investigation. This was carried out using a fixed-effects model with the Mantel–Haenszel method.²⁸ Sensitivity analyses were conducted using a random-effects model with the DerSimonian and Laird method.²⁹

Only one direct head-to-head trial was identified comparing augmentation with AAP with lithium (available as two abstracts and one poster).^{30–32} The main analyses of this trial consisted of people with resistance to either one or two antidepressants in their current episode of depression. This trial also included a mixture of different SSRIs and venlafaxine, a serotonin–norepinephrine reuptake inhibitor (SNRI), and was excluded from the primary analyses. The decision to exclude it from the primary analyses for this review was a result of being unable to obtain subgroup data in the subgroup of patients meeting the inclusion criteria for this review (i.e. patients with resistance to two or more antidepressants and taking SSRIs, with the data reported separately for each SSRI). As a result, it was necessary to carry out an indirect comparison to estimate the efficacy of SSRI + AAP compared with SSRI + lithium. A mixed-treatment comparison (MTC; also called a multiple-treatment meta-analysis and network meta-analysis) was chosen as the method to estimate the effects of SSRI + AAP compared with SSRI + lithium. A MTC can be seen as an extension of traditional pairwise meta-analysis.^{29,33–35}

The MTC was conducted using a fixed- and random-effects model, with the most appropriate model chosen for the reporting of the results. This was determined by the model with the lowest deviance information criterion (DIC).³⁶ DIC measures the fit of the model while penalising for the number of effective

parameters.^{34,37} For the chosen model, the consistency of the evidence was assessed using the posterior mean residual deviance, which should approximate the number of unconstrained data points in a good-fitting model.

For dichotomous outcomes, the odds ratio (OR) is reported as the summary statistic, and, for continuous outcomes, the mean difference (MD).

The primary analysis is:

- response (measured by a reduction of at least 50% in HAMD¹³ or MADRS¹⁴ score).

The secondary analyses are:

- QoL as reported using a validated QoL rating scale²⁵ (e.g. SF-36)
- adverse events (total number of events, individual adverse events for which comparable data were available for both augmentation with AAP and with lithium, and withdrawal rates due to an adverse event)
- withdrawals (all cause) as a surrogate outcome for adherence to medication
- relapse rate
- mortality (all cause).

In addition, remission rates and mean change from baseline MADRS score¹⁴ were also chosen a priori to further assess response to treatment as these were additional clinical parameters that were required for the economic model. Eight-week outcome data were collected where reported. If 8-week data were not available, outcome data from the nearest available time point were collected.

Subgroup analyses were planned in the following populations on only the primary outcome (response), subject to the availability of data:

- different durations of depression (i.e. time since first onset of current episode of depression, short term < 6 months, long term > 6 months)
- class of previous antidepressants (e.g. SSRI or tricyclic antidepressant)
- sex (i.e. male and female)
- age (i.e. those of ≥ 75 years and those of < 75 years old)
- people with different severities of depression, that is, based on trial entry HAMD¹³ rating using the following categories:¹³
 - 8–13 = mild depression
 - 14–18 = moderate depression
 - 19–22 = severe depression
 - ≥ 23 = very severe depression.

In the absence of suitable data to perform a meta-analysis, the available data have been tabulated where possible and discussed in a narrative review.

Heterogeneity

Heterogeneity in pairwise meta-analysis has been explored through consideration of the study populations, methods and interventions, by visualisation of results and, in statistical terms, by the chi-squared test for homogeneity and the I^2 -statistic. Statistically significant heterogeneity has been defined as $p < 0.05$. Levels of inconsistency have been assessed using I^2 and defined as follows. I^2 of: 0–25% = low level of inconsistency; 26–50% = moderate level of inconsistency; and > 50% = high level of inconsistency.³⁸

When statistically significant heterogeneity was detected in any of the primary or secondary analyses, hypothesis-generating subgroup analysis was conducted, although the results from such analyses are highlighted in the text and should be treated with caution. Meta-regression was planned if significant statistical heterogeneity was identified among trials analysed and there were 10 or more trials in the review. However, there were insufficient trials in the review to consider any meta-regression for the pairwise meta-analyses.

For the MTC, where a random-effects model was deemed the best fit, the degree of heterogeneity has been investigated by evaluating the posterior mean tau-squared statistic.³⁹

Sensitivity analysis

The following sensitivity analyses were specified a priori on the primary analysis:

- assuming a 'class' effect with SSRIs and AAPs
- different number of prior antidepressants for the current episode of depression
- changing the quality assessment to include the trial outcomes excluded on grounds of methodological quality, i.e. those categorised as being of high risk of bias
- changing the analysis from using ITT data to PP data.

In addition, the following post hoc sensitivity analysis was conducted:

- limiting the primary analysis to trials reporting response measured by a $\geq 50\%$ reduction on the MADRS scale.¹⁴

None of the trials included in this review was rated as 'high risk of bias' and so the sensitivity analysis including such trials was not required. In addition, no trial reported PP data and so this sensitivity analysis could not be performed.

Publication bias

For each of the primary pairwise meta-analyses, a funnel plot was used to assess publication bias. A regression of normalised effect compared with precision was calculated as a test for small study effects (using $p < 0.10$ to indicate a significant result).^{40,41}

Results

Quantity of research available

The search of electronic databases identified 3717 potentially relevant articles, which, after initial screening, resulted in the identification of 61 potentially relevant full-text articles that were ordered for further screening. An additional three RCTs^{42–44} were identified from the reference list of a systematic review by Wang *et al.*,⁴⁵ and a further RCT⁴⁶ was identified from a systematic review by Nelson *et al.*¹¹ Both systematic reviews were from the 61 full-text articles assessed. In addition, an unpublished poster³² was provided in response to a request for further information on one RCT.³⁰

Following the full assessment of all 66 full-text papers, a total of 11 studies^{30,43,46–53} reported in 15 publications were identified by both reviewers (VH and SB) as meeting the criteria for inclusion: one publication⁵³ included data from two studies, one RCT³⁰ was published as two abstracts^{30,31} with additional data available from an unpublished poster,³² and one RCT⁵¹ was published in four publications^{51,54–56} (one full-text paper and three abstracts). From here on, each RCT will be referred to by only the primary source of the data included in this review, that is, Franco *et al.*³⁰ and Shelton *et al.*⁵¹ It should also be noted that two trials^{49,52} included some patients who had failed to respond to only one antidepressant in their current episode of depression and had a historical failure to a second antidepressant in a previous episode of depression. Another RCT⁴³ may also have included such patients, although the numbers are not reported in

the paper. All three RCTs were included in the primary analyses, although a post hoc decision was taken to perform a sensitivity analysis to assess the impact of excluding these trials from the analysis. In addition, it should be noted that one additional study⁵⁷ that was potentially suitable for inclusion was excluded following appraisal of the full-text papers and after discussion with clinical experts because it was deemed to involve unusually high doses of AAP and the AAP used was one that is not licensed or routinely used in the NHS. The RCT was three armed and compared two different doses of the AAP ziprasidone with placebo augmentation of SSRI. The ziprasidone doses used were 80 mg twice daily (b.i.d.) and 160 mg b.i.d., and the trial was titled a 'pilot study'. In addition, it is noted that there is currently a clinical trial in progress in patients with MDD and failure to respond to an SSRI in their current episode of depression that is assessing the efficacy of augmentation of SSRIs with ziprasidone at doses of 20–80 mg/day compared with augmentation with placebo. This trial is listed on ClinicalTrials.gov as a Phase II trial with an anticipated completion date of March 2013 (ClinicalTrials.gov identifier: NCT00633399).⁵⁸

Of the 11 RCTs agreed by both reviewers (VH and SB) as suitable for inclusion, 10 RCTs were for a comparison of SSRI + AAP with SSRI + placebo/no treatment.^{43,46–53} The remaining study³⁰ compared SSRI or SNRI + AAP with SSRI or SNRI + lithium. Baseline antidepressants in this study were venlafaxine and mixed SSRIs and thus it was agreed among the reviewers (VH, SB and SE) that it should not be included in the primary analysis as the trial author was unable to supply suitable subgroup data. No studies were identified that compared SSRI + lithium with SSRI + placebo/no treatment.

Owing to the absence of suitable trials for the primary analysis including lithium as a comparator, a pragmatic decision was taken to review all the previously screened full-text papers evaluating lithium ($n = 20$), with a view to identifying trials that most closely matched the inclusion criteria. Following this review of previously excluded papers, a single study⁵⁹ that met all of the inclusion criteria, with the exception of the population criterion, was identified. The trial reported in Katona *et al.*⁵⁹ considered the comparative effectiveness of SSRI + lithium with SSRI + placebo in patients who had failed one or more antidepressant regimens. Furthermore, in light of the new NICE guideline for depression in adults (CG90; extracts from this guideline have been reproduced here with permission),¹⁶ it was considered that this trial would suffice as a proxy for a lithium trial in the required population; CG90¹⁶ states that a 'natural cut-off between people who respond to one or two antidepressants compared with those who do not . . . is not supported by the evidence'.

The decision to include this trial in the review was validated by a third reviewer (SJE). In addition, CG90¹⁶ and a systematic review of placebo controlled trials of lithium augmentation therapy in TRD⁶⁰ was used to validate that all of the other potentially relevant SSRI + lithium compared with SSRI + placebo trials had been identified and excluded appropriately. However, the patient population of the surrogate trial for lithium augmentation could be less treatment resistant than the patients in the trials informing treatment augmentation with AAP. The potential impact of this difference in the trial populations is discussed further in the discussion section (see *Assessment of effectiveness*, below).

The search of ClinicalTrials.gov identified no clinical trials for the comparison of SSRI + AAP with SSRI + lithium in the population of interest (i.e. people with TRD and a failure to respond to two or more antidepressants in their current episode of depression) that were completed within the past 12 months or registered as still recruiting patients or ongoing. In addition, clinical experts for this review were not aware of any additional published or unpublished relevant trials.

For a full breakdown of studies included and excluded at each stage of the search and appraisal process, see *Figure 1* (PRISMA diagram). For details of the full-text studies excluded and the individual reasons for exclusion, see *Appendix 4*.

The total number of trials agreed for inclusion in this review was 12 RCTs in 16 publications^{30,43,46–53,59} (two RCTs are reported in one publication⁵³). CG90¹⁶ reported that the DSM-IV tool was used to diagnose depression in most of the evidence reviewed in the guideline and, thus, DSM-IV was the preferred diagnostic

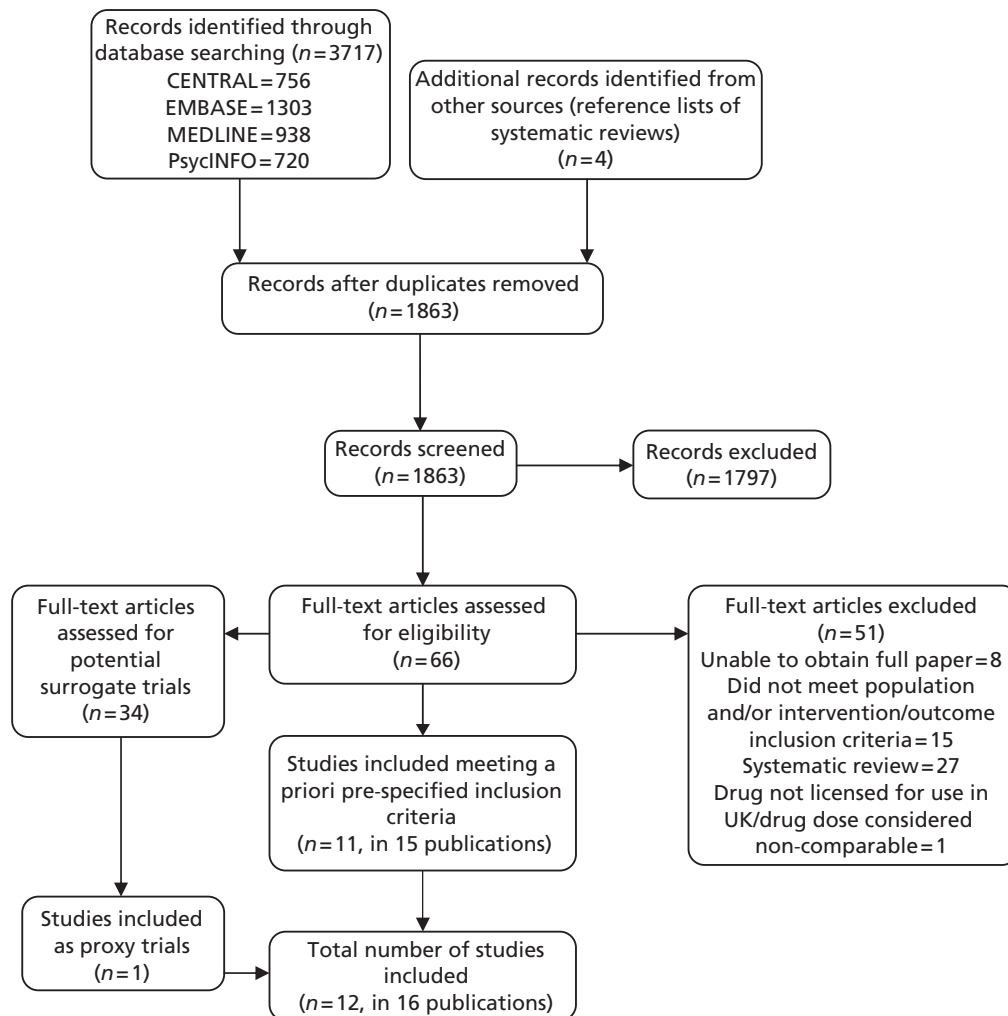


FIGURE 1 Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram for studies included and excluded from the clinical effectiveness review.

measure in the guideline. All but one of the RCTs included in this review required patients to have a diagnosis of MDD based on the DSM-III (one RCT) or DSM-IV (nine RCTs) criteria.

The one RCT⁴³ identified that did not implement DSM criteria to diagnose depression used the *Chinese Classification of Mental Disorders, Version 3 (CCMD-3)* measure. A sensitivity analysis excluding this study is reported in the results section of the report [see *Quality assessment (sensitivity analysis 3)*].

For a summary of the characteristics of each of the studies included in this review, see *Table 2*.

The search of the CCDANCTR-Studies and CCDANCTR-References resulted in the identification of 1487 articles. Initial screening identified three additional potentially relevant papers. All three of these articles represented additional conference abstracts for a study already included following the primary electronic database searches (CENTRAL, EMBASE, MEDLINE and PsycINFO), and the decision was taken not to include them in the results as none of them provided any additional information to that available in the full-text publication.

Quality assessment

All 12 of the included RCTs were assessed for quality using the Cochrane risk-of-bias tool.²⁷ In the overall assessments for each study, as well as the majority of the assessments for the individual outcomes of interest,

TABLE 2 Characteristics of included studies

Author (date)	Study type and setting	Criteria for randomisation to acute-phase treatment	Interventions	No. of patients in each group	Duration of treatment	Additional comments
Berman <i>et al.</i> (2007) ⁴⁷	RCT USA (24 sites)	DSM-IV diagnosis of MDD <i>and</i> history of failure to achieve a response to between one and three antidepressants after at least 6 weeks' treatment at adequate dose and prospective failure to respond to an adequate dose of an investigator-assigned open-label antidepressant (SSRI or venlafaxine) plus single blind (patient blinded) adjunctive placebo during an 8-week screening period	<p>Aripiprazole plus SSRI or venlafaxine</p> <ul style="list-style-type: none"> Aripiprazole was given at an initial dose of 5 mg/day, which was titrated weekly within a range of 2–20 mg/day (maximum 15 mg/day if on fluoxetine or paroxetine) on the basis of response and tolerability Dose of antidepressant received prior to randomisation was given throughout the acute phase of treatment <p><i>Mean dose during final study week (mg/day) (SD):</i> Aripiprazole 11.8</p> <p>Antidepressant doses not reported</p> <p>SSRI or venlafaxine plus placebo</p> <p>Dose of antidepressant received prior to randomisation was given throughout the acute phase of treatment</p> <p><i>Mean modal dose (mg/day) (SD):</i> Doses not reported</p>	Aripiprazole plus SSRI or venlafaxine = 182 patients SSRI or venlafaxine plus placebo = 176 patients	6 weeks	27.4% of patients were on venlafaxine ER Distribution of the SSRI antidepressants in the aripiprazole arm was: <ul style="list-style-type: none"> Escitalopram 33.9% Fluoxetine 17.5% Paroxetine controlled release 7.9% Sertraline 11.9% Distribution of the SSRI antidepressants in the placebo arm was: <ul style="list-style-type: none"> Escitalopram 30.2% Fluoxetine 14.5% Paroxetine controlled release 11.6% Sertraline 17.4%

Author (date)	Study type and setting	Criteria for randomisation to acute-phase treatment	Interventions	No. of patients in each group	Duration of treatment	Additional comments
Berman <i>et al.</i> (2009) ⁴⁸	RCT USA (36 sites)	DSM-IV diagnosis of MDD <i>and</i> history of failure to achieve a response to between one and three antidepressants after at least 6 weeks' treatment at adequate dose and prospective failure to respond to an adequate dose of an open-label antidepressant (SSRI or venlafaxine) plus single blind (patient blinded) adjunctive placebo during an 8-week screening period	<p>Aripiprazole plus SSRI or venlafaxine</p> <ul style="list-style-type: none"> Aripiprazole was given at an initial dose of 5 mg/day, which was titrated weekly within a range of 2–20 mg/day (maximum 15 mg/day if on fluoxetine or paroxetine) on the basis of response and tolerability Dose of antidepressant received prior to randomisation was given throughout the acute phase of treatment <p><i>Mean dose (mg/day) (SD):</i> Aripiprazole 10.7 Antidepressant doses not reported</p> <p>SSRI or venlafaxine plus placebo</p> <p>Dose of antidepressant received prior to randomisation was given throughout the acute phase of treatment</p> <p><i>Mean dose (mg/day) (SD):</i> Placebo 13.9 Antidepressant doses not reported</p>	Aripiprazole plus SSRI or venlafaxine = 177 patients SSRI or venlafaxine plus placebo = 172 patients	6 weeks	In total, 26.2% of patients in the aripiprazole plus SSRI or venlafaxine group were on venlafaxine ER, and 28.8% of patients in the SSRI or venlafaxine plus placebo group were on venlafaxine ER Following randomisation a sex imbalance was noted between the two trial arms, with more females being randomised to the aripiprazole arm (78% vs. 68%) Distribution of the SSRI antidepressants in the trial was: <ul style="list-style-type: none"> Escitalopram 29.6% Fluoxetine 14.2% Paroxetine controlled release 8.9% Sertraline 19.8%

continued

TABLE 2 Characteristics of included studies (continued)

Author (date)	Study type and setting	Criteria for randomisation to acute-phase treatment	Interventions	No. of patients in each group	Duration of treatment	Additional comments
Corya <i>et al.</i> (2006) ⁴⁹	RCT 16 countries (90 sites); no further details reported	DSM-IV diagnosis of MDD <i>and</i> history of failure to achieve a satisfactory response to an SSRI after at least 6 weeks' treatment at a therapeutic dose followed by a prospective partial response (<30% improvement in MADRS total score ⁴⁴) after 7 weeks' open-label treatment with venlafaxine (lead-in phase)	Olanzapine plus fluoxetine: data pooled from four arms, assessing various doses (daily): <ul style="list-style-type: none"> ● Olanzapine 6 mg plus fluoxetine 25 mg ● Olanzapine 12 mg plus fluoxetine 25 mg ● Olanzapine 6 mg plus fluoxetine 50 mg ● Olanzapine 12 mg plus fluoxetine 50 mg Fluoxetine alone <ul style="list-style-type: none"> ● 25 mg or 50 mg alone (daily) <i>Mean modal doses (mg/day; average of all groups):</i> olanzapine = 7.9; fluoxetine = 37.5	Olanzapine plus fluoxetine = 243 patients Fluoxetine alone = 60 patients	12 weeks	Retrospective failure to antidepressant not limited to current episode Eight-armed RCT; specified a priori that data from groups receiving four highest doses of olanzapine and fluoxetine would be pooled for analysis. Other arms in the RCT were a pseudoplacebo (olanzapine 1 mg/day plus fluoxetine 5 mg/day), olanzapine alone, and venlafaxine RCT reports that a subgroup of patients (69.2%) had retrospective failure in their current episode; no data on number of people in each group reported
Feng <i>et al.</i> (2008) ⁴³	RCT China (one site)	CCMD-3 diagnosis of depression <i>and</i> patients had to have previously taken two or more antidepressants with different mechanisms of action in sufficient quantities and for a sufficient duration (each drug treatment time ≥ 6 weeks) with no significant improvement in depressive symptoms	Olanzapine plus fluoxetine <ul style="list-style-type: none"> ● Olanzapine 2.5–5 mg/day ● Fluoxetine 20–40 mg/day Fluoxetine alone <ul style="list-style-type: none"> ● 20–40 mg/day <i>Mean daily doses not reported</i>	Olanzapine plus fluoxetine = 30 patients Fluoxetine alone = 30 patients	8 weeks	Retrospective failures to antidepressant not limited to current episode

Author (date)	Study type and setting	Criteria for randomisation to acute-phase treatment	Interventions	No. of patients in each group	Duration of treatment	Additional comments
Franco <i>et al.</i> (2010) ³⁰	RCT (open label) 12 countries (107 sites)	DSM-IV diagnosis of MDD <i>and</i> history of failure to respond to two antidepressants, no further details reported	Quetiapine ER plus SSRI or venlafaxine <ul style="list-style-type: none"> Quetiapine 300 mg/day Antidepressant doses not reported Lithium plus SSRI or venlafaxine <ul style="list-style-type: none"> Lithium dose 0.6–1.0 µmol/l Antidepressant doses not reported <i>Mean daily doses not reported</i>	Quetiapine plus SSRI or venlafaxine = 114 patients Quetiapine plus SSRI or venlafaxine = 110 patients	6 weeks	Subgroup of a trial (total trial 688 patients) that comprised people with a history of failure to respond to one or more antidepressants Three-armed RCT; third arm assessed the effects of quetiapine alone Of the whole trial population, 36.2% of the quetiapine group was on venlafaxine, and 32.6% of the lithium group was on venlafaxine
Katona <i>et al.</i> (1995) ⁵⁹	RCT UK; number of sites involved not reported	DSM-III-revised diagnosis of MDD <i>and</i> patients were eligible for augmentation with lithium if they had failed to respond to 6 weeks' treatment with an antidepressant (either fluoxetine or lofepramine) during Phase I of the trial Failure to respond defined as: <ul style="list-style-type: none"> reduction in HAMD¹³ of <50% HAMD¹³ at end of Phase I of ≥ 13 	Fluoxetine plus lithium <ul style="list-style-type: none"> Fluoxetine 20 mg Lithium given at an initial dose of 400 mg and increased to 800 mg for days 3–7. Lithium administered to achieve a 12-hour post-dose plasma level of 0.6–1.0 mmol/l Fluoxetine plus placebo <ul style="list-style-type: none"> Fluoxetine 20 mg Lithium carbonate placebo was identical in appearance, taste, and weight to the lithium carbonate tablet <i>Mean daily doses not reported</i>	Fluoxetine 20 mg plus lithium = 17 patients Fluoxetine 20 mg plus placebo = 16 patients	6 weeks	Patients entering augmentation phase were those who had failed to respond to treatment in a double-blind controlled comparison of efficacy and tolerability of fluoxetine and lofepramine RCT assessed augmentation of fluoxetine and of lofepramine with lithium compared with placebo (four-armed RCT)

continued

TABLE 2 Characteristics of included studies (continued)

Author (date)	Study type and setting	Criteria for randomisation to acute-phase treatment	Interventions	No. of patients in each group	Duration of treatment	Additional comments
Marcus <i>et al.</i> (2008) ⁵⁰	RCT USA (36 sites)	DSM-IV diagnosis of MDD and history of failure to achieve a response to between one and three antidepressants after at least 6 weeks' treatment at adequate dose and prospective failure to respond to an adequate dose of an open-label investigator-assigned antidepressant (SSRI or venlafaxine) plus single-blind (patient-blinded) placebo during an 8-week screening period	<p>Aripiprazole plus SSRI or venlafaxine</p> <ul style="list-style-type: none"> Aripiprazole was given at an initial dose of 5 mg/day, which was titrated weekly within a range of 2–20 mg/day (maximum 15 mg/day if on fluoxetine or paroxetine) on the basis of response and tolerability Dose of antidepressant received prior to randomisation was given throughout the acute phase of treatment <p><i>Mean dose in final week (mg/day) (SD):</i> Aripiprazole 11.0 Antidepressant doses not reported</p> <p>SSRI or venlafaxine plus placebo</p> <ul style="list-style-type: none"> Dose of antidepressant received prior to randomisation was given throughout the acute phase of treatment <p><i>Mean dose in final week (mg/day) (SD):</i> Placebo 15.3 Antidepressant doses not reported</p>	Aripiprazole plus SSRI or venlafaxine = 191 patients SSRI or venlafaxine plus placebo = 190 patients	6 weeks	28.1% of patients were on venlafaxine Distribution of the SSRI antidepressants in the trial was: <ul style="list-style-type: none"> Escitalopram 30.4% Fluoxetine 14.7% Paroxetine controlled release 7.1% Sertraline 19.7%

Author (date)	Study type and setting	Criteria for randomisation to acute-phase treatment	Interventions	No. of patients in each group	Duration of treatment	Additional comments
Mattingly <i>et al.</i> (2006) ⁴⁶	RCT USA (no further details reported)	DSM-IV diagnosis of MDD <i>and</i> history of at least one failure to achieve a satisfactory response to an antidepressant after at least 4 weeks' treatment at a clinically appropriate dose within the current episode of MDD, and a further treatment failure defined as HAM-D-17 rating of ≥ 20 after at least 6 weeks' treatment with an SSRI or SNRI, which must have been different from the earlier antidepressant failed	Quetiapine plus SSRI or SNRI <ul style="list-style-type: none"> Quetiapine titrated from 50 mg to 200–400 mg/day plus ongoing treatment with an SSRI or SNRI <p><i>Mean final dose (mg/day) (SD):</i> quetiapine = 268 (71.1); SSRI/SNRI = NR</p> <p>SSRI or SNRI plus placebo</p> <ul style="list-style-type: none"> Ongoing treatment with an SSRI or SNRI plus placebo <p><i>Mean final dose (mg/day) (SD):</i> Placebo = 341 (53.9) SSRI/SNRI = NR</p>	Quetiapine plus SSRI or SNRI = 26 patients SSRI or SNRI plus placebo = 14 patients	8 weeks	Two patients (one in each treatment arm) were on both an SSRI and an SNRI. A further nine patients (six in quetiapine group and three in the placebo group) were on SNRIs (not further defined)
Shelton <i>et al.</i> (2001) ⁵¹	RCT USA (two sites)	DSM-IV diagnosis of MDD <i>and</i> history of failure to achieve a response to an antidepressant other than an SSRI after at least 4 weeks' treatment and prospective failure to respond to an adequate dose of fluoxetine (SSRI) during a 6-week screening period	Olanzapine plus fluoxetine <ul style="list-style-type: none"> Olanzapine was given at an initial dose of 5 mg/day, which was titrated weekly within a range of 5–20 mg/day on the basis of response and tolerability Dose of fluoxetine received 1 week before randomisation was given throughout the acute phase of treatment <p><i>Mean modal dose (mg/day) (SD):</i> olanzapine = 13.5 (4.1); fluoxetine = 52.0 (10.3)</p>	Olanzapine plus fluoxetine = 10 patients Fluoxetine plus placebo = 10 patients	8 weeks	Three-armed RCT; third arm assessed the effects of olanzapine alone

continued

TABLE 2 Characteristics of included studies (continued)

Author (date)	Study type and setting	Criteria for randomisation to acute-phase treatment	Interventions	No. of patients in each group	Duration of treatment	Additional comments
Shelton <i>et al.</i> (2005) ⁵²	RCT USA and Canada (71 sites)	DSM-IV diagnosis of MDD and history of at least one failure to achieve a satisfactory response to an SSRI after at least 4 weeks' treatment at a therapeutic dose occurring within either the current episode or a historical episode of MDD, and prospective treatment failure to 7 weeks' treatment with nortriptyline (Allegron [®] , King) (open-label, dose-escalation phase); prospective failure defined as <30% improvement (decrease) in MADRS total score ¹⁴ from baseline	<p>Fluoxetine plus placebo</p> <ul style="list-style-type: none"> Dose of fluoxetine received 1 week before randomisation was given throughout the acute phase of treatment <p>Mean modal dose (mg/day) (SD): 52.0 (14.0)</p>			
			<p>Olanzapine plus fluoxetine</p> <ul style="list-style-type: none"> Patients could receive either olanzapine 6 mg/day plus fluoxetine 25 mg/day or olanzapine 12 mg/day plus fluoxetine 50 mg/day Olanzapine was started at dose of 6 mg/day and could be increased to 12 mg/day at clinician's discretion Fluoxetine was started at dose of 25 mg/day and could be increased to 50 mg/day at clinician's discretion <p>Mean modal dose (mg/day) (SD): olanzapine = 8.5 (3.1); fluoxetine = 35.6 (12.7)</p> <p>Fluoxetine plus placebo</p> <ul style="list-style-type: none"> Fluoxetine was started at dose of 25 mg/day and could be increased to 50 mg/day at clinician's discretion <p>Mean modal dose (mg/day) (SD): 35.8 (12.8)</p>	Olanzapine plus fluoxetine = 146 patients Fluoxetine plus placebo = 142 patients	8 weeks	Retrospective failure to antidepressant not limited to current episode RCT reports that a subgroup of patients (62.8%) had retrospective failure in their current episode; no data on number of people in each group reported Four-armed RCT; the remaining arms assessed the effects of olanzapine alone and nortriptyline

Author (date)	Study type and setting	Criteria for randomisation to acute-phase treatment	Interventions	No. of patients in each group	Duration of treatment	Additional comments
Thase <i>et al.</i> (2007), ⁵³ studies 'a' and 'b'	RCTs USA and Canada; no. of sites involved not reported Two separate identical 8-week, double-blind, parallel-group RCTs reported in the same publication	DSM-IV diagnosis of MDD <i>and</i> history of a failure to achieve a satisfactory response to an antidepressant (except fluoxetine) after at least 6 weeks' treatment at a therapeutic dose within the current episode of MDD (based on the investigators' clinical judgement) and prospective failure to 8 weeks' open-label treatment with fluoxetine (lead-in phase); patients were excluded from the lead-in phase if they showed evidence of psychotic features or response to fluoxetine ($\geq 25\%$ decrease in the IVR HAMD-17 score or an IVR HAMD-17 score of < 18 or a $> 15\%$ decrease between weeks 7 and 8 of the lead-in phase)	Olanzapine plus fluoxetine <ul style="list-style-type: none"> Olanzapine 6, 12 or 18 mg/day: patients were started at 6 mg/day and were required to be titrated to the higher dose unless there were issues with safety or tolerability Fluoxetine 50 mg/day <p><i>Mean modal dose (mg/day) (SD):</i> olanzapine = 8.6 (4.7); fluoxetine = 48.8 (7.8)</p> Fluoxetine alone <ul style="list-style-type: none"> 50 mg/day <p><i>Mean modal dose (mg/day) (SD):</i> 49.5 (4.9)</p> <p>Fluoxetine alone = 102 patients</p>	Study 1: Olanzapine plus fluoxetine = 102 patients Fluoxetine alone = 104 patients Study 2: Olanzapine plus fluoxetine = 98 patients	8 weeks	Three-armed RCT; third arm assessed the effects of olanzapine alone

ER, extended release; HAMD-17, the original version of the HAMD, which was based on 17 items; IVR, interactive voice response; NR, not reported; SD, standard deviation.

all of the trials were rated as unclear risk of bias. This was generally the result of a lack of information being reported in the methods, and that, despite contacting study authors, the additional information could not be obtained. This reason for rating studies as having unclear risk of bias is not unusual as it has been reported elsewhere that unclear risk is likely to be assigned owing to poor reporting of how a trial was conducted rather than a poorly conducted trial.⁶¹ For full details of the risk-of-bias assessments for each study see *Appendix 5*.

Assessment of effectiveness

The RCTs meeting the inclusion criteria for the primary analyses in the clinical effectiveness review comprise trials comparing SSRI + AAP with SSRI + placebo/no treatment, and SSRI + lithium with SSRI + placebo. These trials were used to create a network for the MTC to address the review question regarding comparison of the clinical effectiveness of SSRI + AAP with that of SSRI + lithium. The individual clinical effectiveness results are presented separately below for each of the following comparisons:

1. SSRI + AAP vs. SSRI + placebo/no treatment
2. SSRI + lithium vs. SSRI + placebo
3. SSRI + AAP vs. SSRI + lithium.

Selective serotonin reuptake inhibitor plus atypical antipsychotic compared with selective serotonin reuptake inhibitor plus placebo/no treatment

A total of six trials were identified that met the criteria for inclusion in the primary analyses.^{43,49,51–53} All six RCTs compared fluoxetine (SSRI) + olanzapine (AAP) with fluoxetine (SSRI) alone (or fluoxetine + placebo). Corya *et al.*⁴⁹ and Shelton *et al.*^{51,52} were reported as using an AAP placebo tablet, and the two studies reported in Thase *et al.*⁵³ were double blind, which suggests that an AAP placebo tablet was used. Feng *et al.*⁴³ also reported limited information on the conduct of the trial: there was no mention of blinding and so it is unlikely that a placebo tablet was provided to the fluoxetine-alone treatment group. For simplicity, from here onwards, SSRI + placebo/no treatment will be referred to as 'SSRI alone'. It should also be noted that the Thase *et al.* studies⁵³ were two identical concurrent studies that were reported in a single publication. For the purpose of this review, many of the results in the ITT population are limited to a pooled analysis of these two studies⁵³ because the appropriate data for the individual studies could not be obtained. However, the use of pooled data in the analyses is highlighted in the corresponding text.

A further four trials^{46–48,50} were included in the class-based sensitivity analysis because they allowed a range of antidepressants, including SNRIs such as venlafaxine, as the baseline for augmentation and they did not present individual subgroup results for each SSRI. Three of these trials^{47,48,50} compared an antidepressant + aripiprazole with antidepressant + placebo and the remaining trial⁴⁶ compared an antidepressant + quetiapine with antidepressant + placebo. The antidepressants included various SSRIs and SNRIs, although in all three of the aripiprazole trials the SNRI was limited to venlafaxine. All four trials were included in a sensitivity analysis to assess the efficacy of augmentation if AAPs and SSRIs are assumed to have a class-based effect, rather than assuming that different drugs within each class have different efficacy. As over 70% of patients in each trial received an SSRI as baseline therapy it was agreed by the reviewers (SB, SJE and VH) to include the four trials^{46–48,50} in the sensitivity analysis. Subgroup data for those on SSRI alone as baseline therapy were also sought from corresponding authors. These SSRI subgroup data were provided for two of the trials^{47,50} for the outcome of mean change in MADRS score¹⁴ from baseline at study end point from a pooled analysis.⁶² Other corresponding authors either did not reply or were unable to provide data on this subgroup.

Results for selective serotonin reuptake inhibitor plus atypical antipsychotic compared with selective serotonin reuptake inhibitor alone

Response This was predefined in the protocol as a reduction of $\geq 50\%$ in MADRS¹⁴ or HAMD¹³ score from baseline at the trial end point. It was reported in all six trials^{43,49,51–53} that met the inclusion criteria for this comparison, although the data for Thase *et al.*⁵³ are reported as a pooled analysis. Five of the trials^{49,51–53}

reported response based on the MADRS scale¹⁴ and the remaining trial⁴³ used the HAMD scale. The results of the meta-analysis (fixed effects) demonstrated a statistically significant benefit of fluoxetine + olanzapine over fluoxetine alone [OR 1.48; 95% confidence interval (CI) 1.13 to 1.94] with a moderate level of statistical heterogeneity ($I^2 = 53%$; $p = 0.07$) (Figure 2).

Visual inspection of the funnel plot did not appear to infer publication bias (Figure 3). A regression of normalised effect compared with precision (otherwise known as 'Egger's regression test' or more simply the 'Egger's test') was calculated as a test for small study effects (using $p < 0.10$ as an indicator of a significant result).⁴⁰ The Egger's test was not statistically significant ($p = 0.17$).

A random-effects model was used in a sensitivity analysis to explore whether or not the results were sensitive to the choice of using a fixed-effects model in the primary analysis. The random effects meta-analysis remained statistically significant in favour of treatment with fluoxetine + olanzapine (OR 1.60; 95% CI 1.01 to 2.53). These results suggest that augmentation of fluoxetine with olanzapine significantly improves the likelihood of treatment response in patients with TRD compared with those receiving treatment with fluoxetine (SSRI) alone.

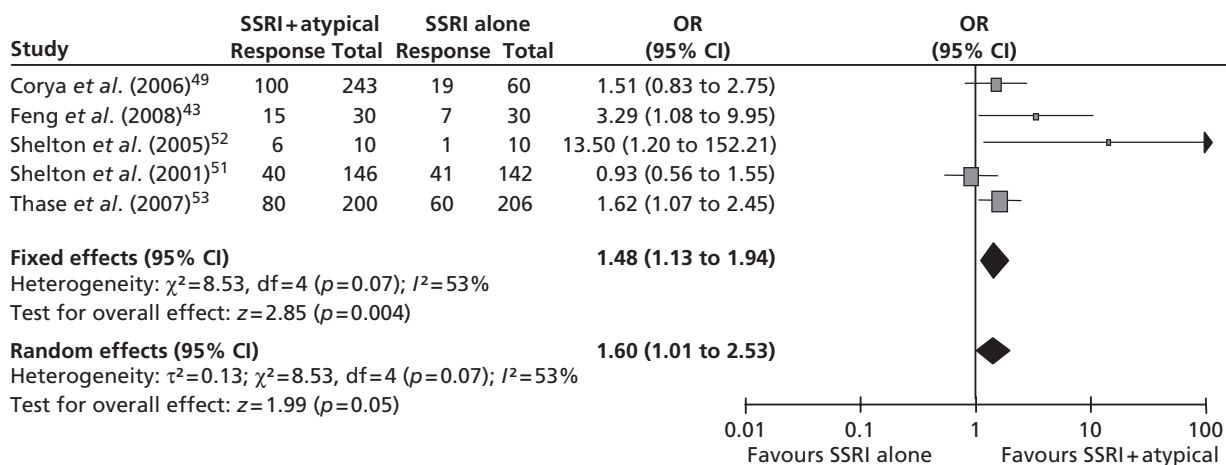


FIGURE 2 Results of meta-analysis for $\geq 50%$ response comparing SSRI + AAP with SSRI alone. df, degrees of freedom.

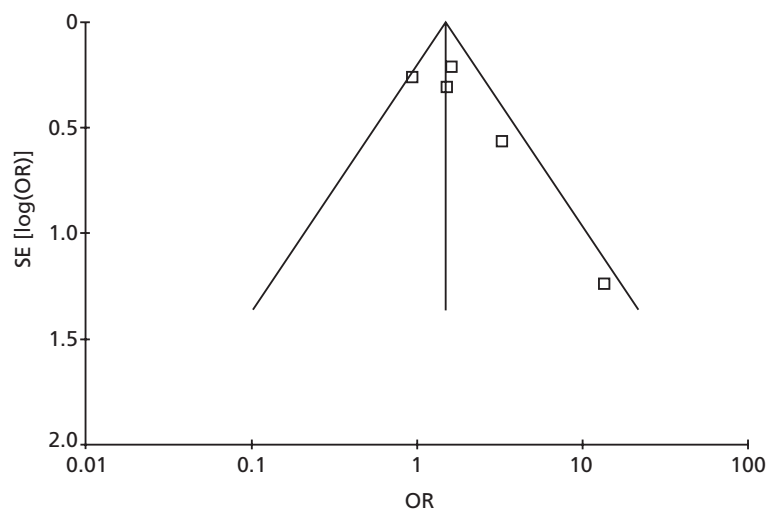


FIGURE 3 Funnel plot for response. SE, standard error.

Mean change in MADRS score This analysis was conducted to provide clinical data for use in the economic model (see *Chapter 5, Acute treatment phase*). Four trials were included in the meta-analysis.^{49,52,53} Using the least square MD from baseline in MADRS score at study end point resulted in a statistically significant MD of -2.04 (95% CI -3.25 to -0.83) in favour of fluoxetine + olanzapine. This equates to a mean reduction of two points on the MADRS scale with olanzapine augmentation of fluoxetine compared with fluoxetine alone. However, there was a high level of heterogeneity that was statistically significant ($I^2 = 73\%$; $p = 0.01$). One possible explanation for the heterogeneity could be related to the study population of Thase *et al.* (study b).⁵³ It reported a much larger MD (-5.9) between fluoxetine + olanzapine and fluoxetine alone compared with the other trials included in the meta-analysis (values between -0.2 and -2.36). Similarly, Thase *et al.* (study b)⁵³ was the only trial to report a statistically significant difference in mean change in MADRS score ($p < 0.001$). Random effects meta-analysis also resulted in a statistically significant MD in favour of fluoxetine + olanzapine compared with fluoxetine alone (MD -2.40 ; 95% CI -4.76 to -0.04) (*Figure 4*).

An exploratory meta-analysis was conducted to assess the impact of removing Thase *et al.* (study b).⁵³ This resulted in the removal of the significant statistical heterogeneity ($I^2 = 0\%$; $p = 0.38$) and the statistically significant effect on MD -1.15 (95% CI -2.49 to 0.19), although the trend was still in favour of fluoxetine + olanzapine compared with fluoxetine alone.

Visual inspection of the funnel plot did not appear to indicate publication bias and the Egger's test was not statistically significant ($p = 0.23$).

Remission In total five RCTs^{43,49,52,53} reported remission, although the data for Thase *et al.*⁵³ are reported as a pooled analysis. It was noted that the criteria for remission varied slightly between the RCTs. In Corya *et al.*⁴⁹ and Shelton *et al.*,⁵² the definition of remission used was two consecutive MADRS total scores of ≤ 8 , whereas Thase *et al.*⁵³ used a definition of a MADRS total score of ≤ 10 at the study end point. Feng *et al.*⁴³ used the HAMD and their definition of remission was a HAMD score of < 7 .

Clinical experts for this review consider that the most commonly used definition of remission in patients with TRD is a MADRS total score of ≤ 10 ; there is some uncertainty regarding the equivalent definition for remission using the HAMD rating scale, although the clinical experts for this report consider that a score of ≤ 7 is commonly used. The results for all five trials^{43,49,52,53} were analysed as reported in the RCTs with no attempt made to convert them to a standard definition of remission. The results of the fixed-effects meta-analysis demonstrated a statistically significant increase in remissions in patients treated with olanzapine + fluoxetine compared with fluoxetine alone (OR 1.77; 95% CI 1.27 to 2.47), with no statistical heterogeneity ($I^2 = 0\%$; $p = 0.75$) (*Figure 5*). There was no evidence of publication bias in the funnel plot and the Egger's test was not statistically significant ($p = 0.61$).

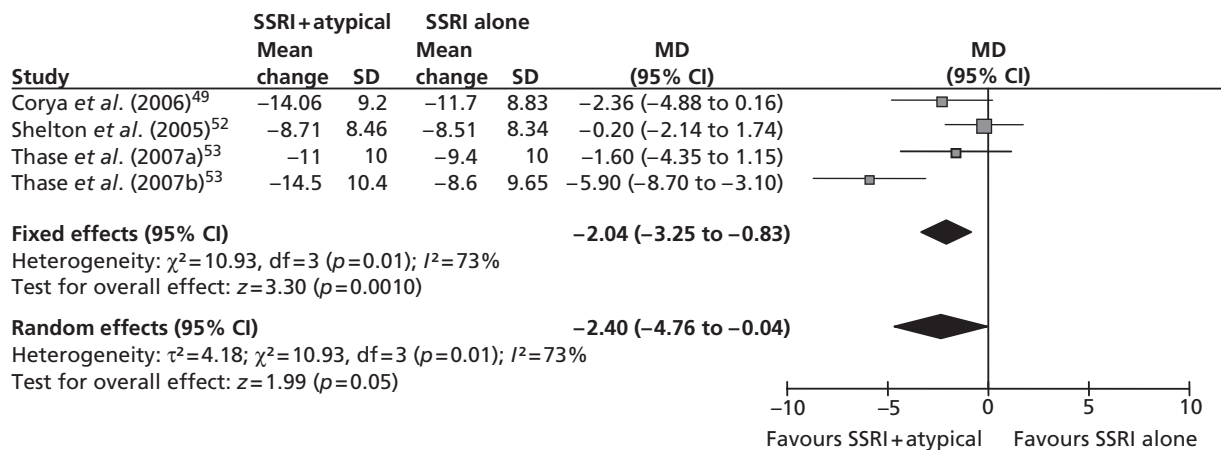


FIGURE 4 Results of meta-analysis for mean change in MADRS score (SSRI + AAP vs. SSRI alone). *df*, degrees of freedom; *SD*, standard deviation.

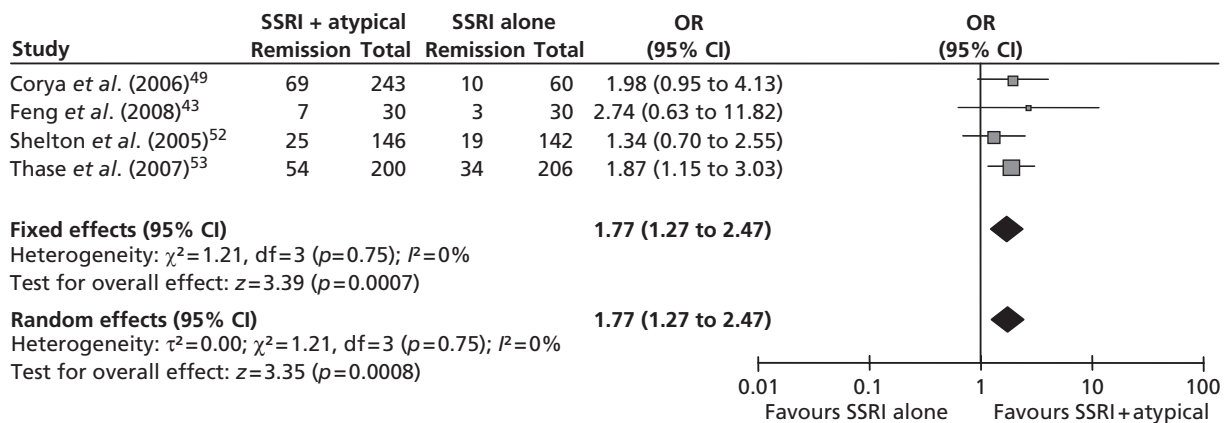


FIGURE 5 Results of meta-analysis for remission comparing SSRI + AAP vs. SSRI alone. *df*, degrees of freedom.

Quality of life QoL data were only reported in one paper as a pooled analysis for the two Thase *et al.*⁵³ RCTs. The data were limited to changes on the Sheehan disability scale and the SF-36 scale, and are summarised in *Table 3*. In the QoL domains assessed, patients receiving fluoxetine + olanzapine generally showed greater improvements than with fluoxetine-alone patients, although the difference in the SF-36 mental health summary score was not statistically significant ($p > 0.05$).

Withdrawals (all cause) Data on all-cause withdrawals were reported in five trials,^{49,51–53} although the data for the two Thase *et al.* studies⁵³ were reported as a pooled analysis. The results of the fixed-effects meta-analysis suggest a statistically non-significant reduction in discontinuations with fluoxetine alone compared with fluoxetine + olanzapine (OR 1.25; 95% CI 0.91 to 1.71), with no statistical heterogeneity ($I^2 = 0\%$; $p = 0.51$) (*Figure 6*). Visual inspection of the funnel plot did not appear to indicate publication bias and the Egger's test was not statistically significant ($p = 0.13$).

Withdrawals due to adverse events A total of five studies^{43,49,52,53} were suitable for inclusion in this meta-analysis. Thase *et al.*⁵³ reported results only for a pooled analysis of the two studies. The results of the meta-analysis show a statistically significantly lower risk of withdrawal due to an adverse event with fluoxetine alone than with fluoxetine + olanzapine (OR 3.85; 95% CI 2.03 to 7.29), with no statistical heterogeneity ($I^2 = 0\%$; $p = 0.40$) (*Figure 7*). The funnel plot and the Egger's test were deemed inappropriate for use for this outcome as the meta-analysis contained fewer than four trials.

Adverse events A diverse range of adverse events were reported across the trials included in this review. As it seemed likely that a substantial analytical effort would be rewarded with little gain, a pragmatic

TABLE 3 Quality-of-life results for Thase *et al.*⁵³ pooled analysis

Quality-of-life measure	Fluoxetine + olanzapine: <i>n</i> (SD)	Fluoxetine alone: <i>n</i> (SD)	<i>p</i> -value
Mean improvement from baseline on Sheehan Disability Scale – leisure item	– 1.6 (2.8)	– 1.1 (2.6)	0.027
Mean improvement from baseline on Sheehan Disability Scale – family item	– 1.7 (2.7)	– 1.2 (2.6)	0.047
Mean change from baseline on SF-36 summary mental score	8.9 (12.6)	7.3 (12.3)	0.175
Mean change from baseline on SF-36 summary physical score	2.1 (9.0)	0.4 (8.7)	0.028

SD, standard deviation.

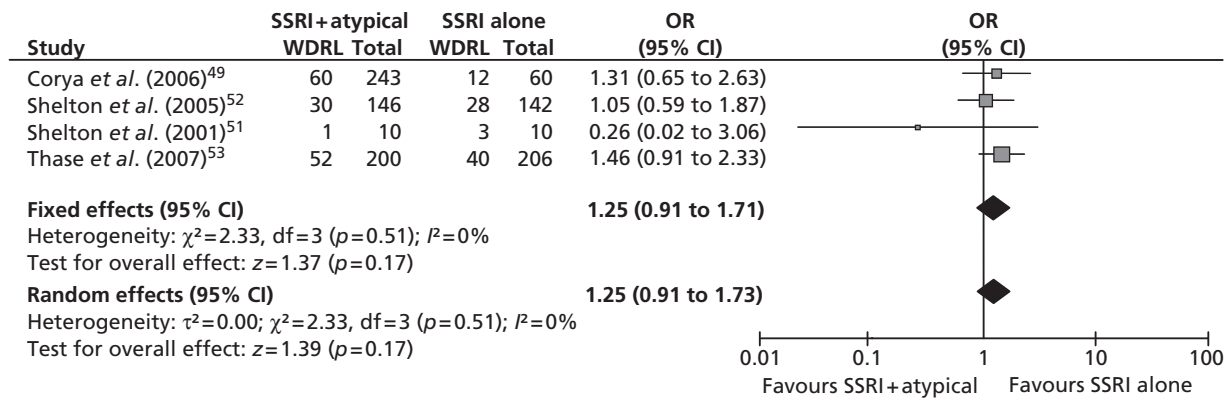


FIGURE 6 Results of meta-analysis for all-cause withdrawals comparing SSRI + AAP vs. SSRI alone. df, degrees of freedom; WDRL, number of patients withdrawing from the study.

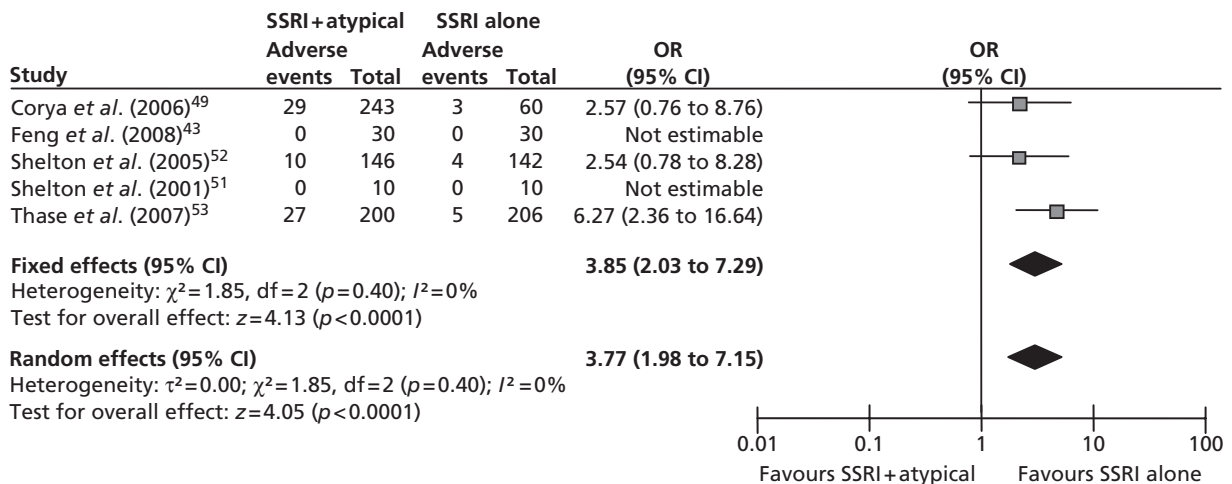


FIGURE 7 Results of meta-analysis for withdrawals due to adverse events comparing SSRI + AAP vs. SSRI alone. df, degrees of freedom.

decision was taken to analyse only the total number of people reporting adverse events and the individual adverse events for which there were comparable data for both adjuvant treatment regimens. The results with lithium and from the indirect comparison are presented later in this report. Only two RCTs^{43,52} informed the comparison of fluoxetine + olanzapine with olanzapine alone. The result of a meta-analysis of the two trials was not statistically significant but did suggest a trend favouring treatment with fluoxetine alone compared with fluoxetine + olanzapine (OR 1.60; 95% CI 0.91 to 2.83), with no statistical heterogeneity ($I^2 = 0\%$; $p = 0.39$).

The individual adverse events with sufficient data for analysis were tremor, somnolence, increased appetite, dry mouth and headache. The individual adverse events analyses generally resulted in statistically significant results when the fixed-effects model was used (in favour of fluoxetine alone, with the exception of headache, which favoured fluoxetine + olanzapine) and statistically non-significant results with wide 95% CIs when a random-effects model was used. The meta-analyses had high levels of heterogeneity that was statistically significant. The results for the individual adverse events should thus be interpreted with caution. For full details of the adverse events results, see *Figure 8*.

The funnel plot and the Egger's test were deemed inappropriate for use for any of the adverse event outcomes as no meta-analysis contained four or more trials.

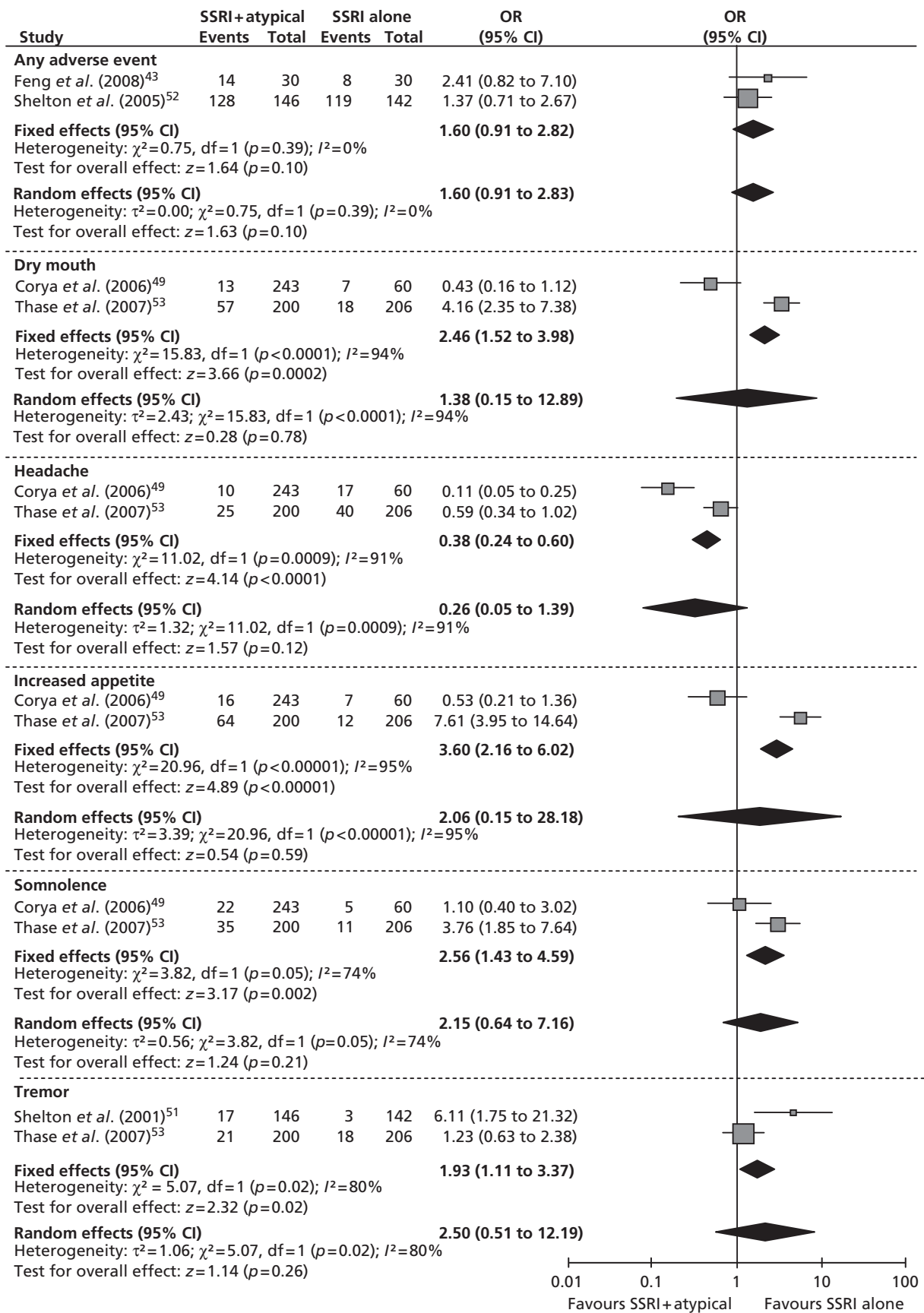


FIGURE 8 Results of meta-analysis for adverse events (SSRI + AAP vs. SSRI alone). *df*, degrees of freedom.

In addition to the adverse events reported here, there were also data from one trial⁴⁹ on the adverse event of dizziness, for which there are similar data for lithium, which are presented later in this report. The Corya *et al.* trial⁴⁹ reported an OR of 0.31 (95% CI 0.13 to 0.73), suggesting a statistically significant reduction in dizziness with fluoxetine + olanzapine compared with fluoxetine alone. The total number of patients receiving fluoxetine alone was small in comparison with the number receiving fluoxetine + olanzapine (60 vs. 243, respectively).

Relapse rate This outcome was not reported in any of the RCTs suitable for inclusion in this analysis.

Mortality This outcome was not reported in any of the RCTs suitable for inclusion in this analysis.

Sensitivity analysis assuming class response The sensitivity analysis, assuming that SSRIs have a class effect and all AAPs have a class effect, was conducted on the primary outcome of response (*Figure 9*). Response was chosen as it was expected a priori to be the outcome most commonly reported by trials. The meta-analysis for this outcome included all 10 trials identified for the comparison of SSRI + AAP with SSRI alone.^{43,46-53} The results demonstrate a statistically significant improved response with SSRI + AAP compared with SSRI alone when either the fixed or random-effects models are used (fixed effects, OR 1.78; 95% CI 1.48 to 2.15; random effects, OR 1.83; 95% CI 1.38 to 2.42). This is consistent with the primary analysis for response, where no class effect is assumed (fixed effects, OR 1.48; 95% CI 1.13 to 1.94).

This meta-analysis includes some patients receiving SNRI as a baseline treatment to be augmented with an AAP. The overall impact of this on the results is unclear as any bias introduced would equally affect both trial arms within a RCT. The funnel plot for this analysis appeared to be symmetrical and the Egger's test was not statistically significant ($p=0.14$).

Selective serotonin reuptake inhibitor plus lithium compared with selective serotonin reuptake inhibitor plus placebo

Only one RCT⁵⁹ was identified for inclusion in the analyses of SSRI + lithium compared with SSRI alone. As highlighted earlier, it should be noted that this trial is a surrogate trial used because of an absence of lithium trials in the required population. The patient population of Katona *et al.*⁵⁹ had failed to respond to one or more antidepressants in the current episode of depression, rather than two or more antidepressants as specified by the population inclusion criterion. The use of a proxy trial facilitated an indirect comparison between augmentation of SSRI with lithium and augmentation of SSRI with an AAP. However, patients within this trial may be less treatment resistant compared with the patients in the trials informing treatment

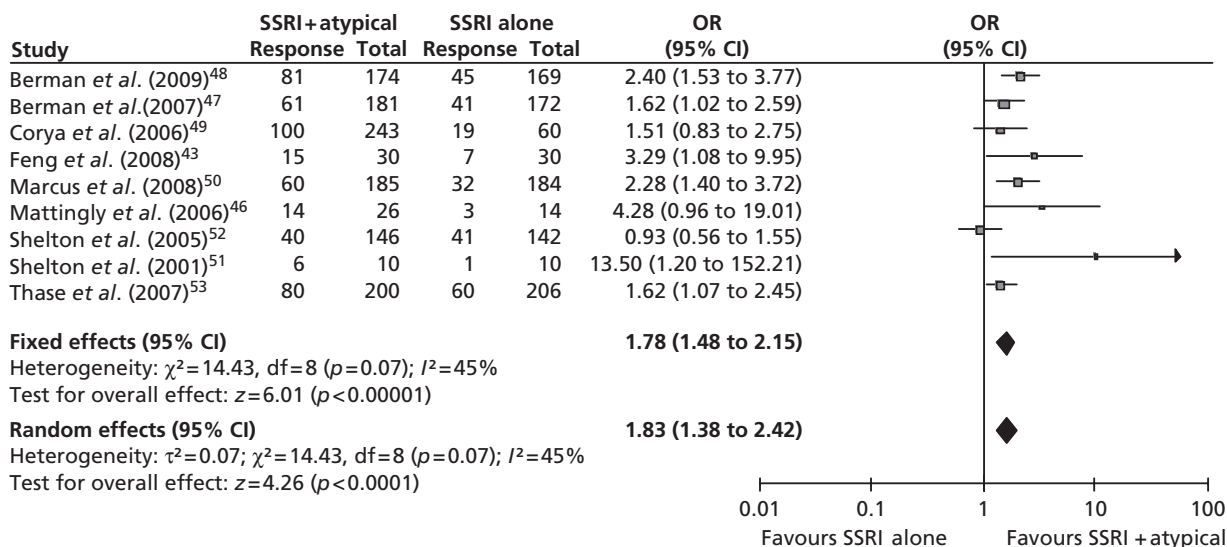


FIGURE 9 Results of sensitivity meta-analysis for response (SSRI + AAP vs. SSRI alone). *df*, degrees of freedom.

augmentation with AAP. The potential impact of this difference in the trial populations is discussed further in the discussion section (see *Assessment of effectiveness*, below).

Results for selective serotonin reuptake inhibitor plus lithium compared with selective serotonin reuptake inhibitor alone

Response Two different definitions of treatment response were presented in the report for Katona *et al.*,⁵⁹ neither of which directly met the review criteria of response (i.e. a $\geq 50\%$ reduction in MADRS or HAMD score from baseline score) (Box 4).

The post hoc definition for response was chosen by the trial authors to enable the data from the trial to be compared with that of other similar trials in a meta-analysis that was conducted by the trial author and also reported in Katona *et al.*⁵⁹

The result of the trial using the a priori definition for response demonstrated a non-significant trend in favour of fluoxetine + lithium compared with fluoxetine alone (OR 1.48; 95% CI 0.37 to 5.95). Using the post hoc criteria for response there is a larger treatment effect that favours fluoxetine + lithium compared with fluoxetine alone but the difference between treatment effects remains non-significant (OR 3.85; 95% CI 0.80 to 18.62).

Mean change in MADRS score The baseline MADRS score in Katona *et al.*⁵⁹ was 26.06 [standard deviation (SD) 4.93] in the fluoxetine + lithium group, and 26.25 (SD 5.29) in the fluoxetine-alone group. The MADRS scores at week 6 of augmentation therapy were 15.88 (SD 10.27) in the fluoxetine + lithium group, and 19.45 (SD 7.39) in the fluoxetine + placebo group. The between-group mean change from baseline MADRS score at week 6 was -3.79 (fluoxetine + lithium vs. fluoxetine alone; 95% CI -11.25 to 3.67) and was statistically non-significant.

Remission Data for this outcome were not reported in the published RCT and were not available from the trial author or sponsor.

Quality of life Data for this outcome were not reported in the published clinical trial and were not available from the trial author or sponsor.

BOX 4 Definitions of response used in Katona *et al.*⁵⁹

Definition 1 (a priori definition)

Reduction in HAMD score of at least 50% from Phase I baseline (i.e. prior to a 6-week lead-in phase of fluoxetine treatment).

Reduction in HAMD score of at least 25% from Phase II baseline (i.e. prior to commencement of lithium or placebo augmentation therapy).

Final HAMD score of < 13 .

Definition 2 (post hoc definition)

Reduction in HAMD score of at least 50% from Phase I baseline (i.e. prior to a 6-week lead-in phase of fluoxetine treatment).

Reduction in HAMD score of at least 25% from Phase II baseline (i.e. prior to commencement of lithium or placebo augmentation therapy).

Final HAMD score of < 10 .

Withdrawals (all cause) Discontinuations from the Katona *et al.*⁵⁹ trial were slightly lower with fluoxetine + lithium (4/17) compared with fluoxetine alone (5/16). This difference was not statistically significant (OR 0.68; 95% CI 0.15 to 3.16). The reasons for withdrawals in the lithium group were patient's decision ($n = 1$), physician's decision ($n = 2$) and protocol violation ($n = 1$). The reasons for withdrawal with fluoxetine alone were lack of efficacy ($n = 1$), patient's decision ($n = 2$), protocol violation ($n = 1$), and adverse event ($n = 1$).

Withdrawals due to adverse events As already described above, there was only one withdrawal due to an adverse event reported in Katona *et al.*⁵⁹ This was in the fluoxetine-alone group. No further details were provided in the additional information received from the study sponsor for this outcome.

Adverse events There was a non-statistically significant difference in favour of treatment with fluoxetine-alone for the total number of adverse events compared with the fluoxetine + lithium group (OR 1.95; 95% CI 0.43 to 8.83).

The individual adverse event data for the outcomes of interest (as discussed above in the SSRI + AAP vs. SSRI alone comparison) are presented in *Table 4*. None of the results shows a statistically significant difference between treatment groups. However, it should be noted that these are not necessarily the adverse events that would be commonly expected with treatment with lithium or fluoxetine; these will be discussed further in the discussion section (see *Assessment of effectiveness*, below). In addition, as a result of the small number of people in the Katona *et al.*⁵⁹ trial, the number of people reporting any single adverse event is relatively small and the study was not powered to detect between-group differences in adverse events. The results presented should thus be interpreted with caution.

Relapse rate Data for this outcome were not reported in the published clinical trial and were not available from the trial author or sponsor.

Mortality Data for this outcome were not reported in the published clinical trial and were not available from the trial author or sponsor.

Sensitivity analysis assuming class response Analysis not required, as no additional trials were identified as suitable for inclusion in this comparison.

Selective serotonin reuptake inhibitor plus atypical antipsychotic compared with selective serotonin reuptake inhibitor plus lithium

Seven trials^{43,49,51–53,59} were identified that met the criteria for inclusion in the network meta-analysis (MTC). Of these trials, six RCTs^{43,49,51–53} compared fluoxetine + olanzapine with fluoxetine alone, and one trial compared fluoxetine + lithium with fluoxetine alone.⁵⁹ See *Figure 10* for network diagram.

TABLE 4 Adverse event results reported in Katona *et al.*⁵⁹

Adverse event	Fluoxetine + lithium		Fluoxetine alone		OR (95% CI)
	<i>n</i>	<i>N</i>	<i>n</i>	<i>N</i>	
Dizziness	0	17	1	16	0.30 (0.01 to 7.79)
Dry mouth	2	17	1	16	0.44 (0.04 to 5.36)
Headache	3	17	2	16	0.58 (0.08 to 4.01)
Increased appetite	1	17	0	16	0.30 (0.01 to 7.79)
Somnolence	1	17	2	16	2.00 (0.16 to 24.48)
Tremor	1	17	0	16	0.30 (0.01 to 7.79)
Total no. of people experiencing an adverse event	13	17	10	16	1.95 (0.43 to 8.83)

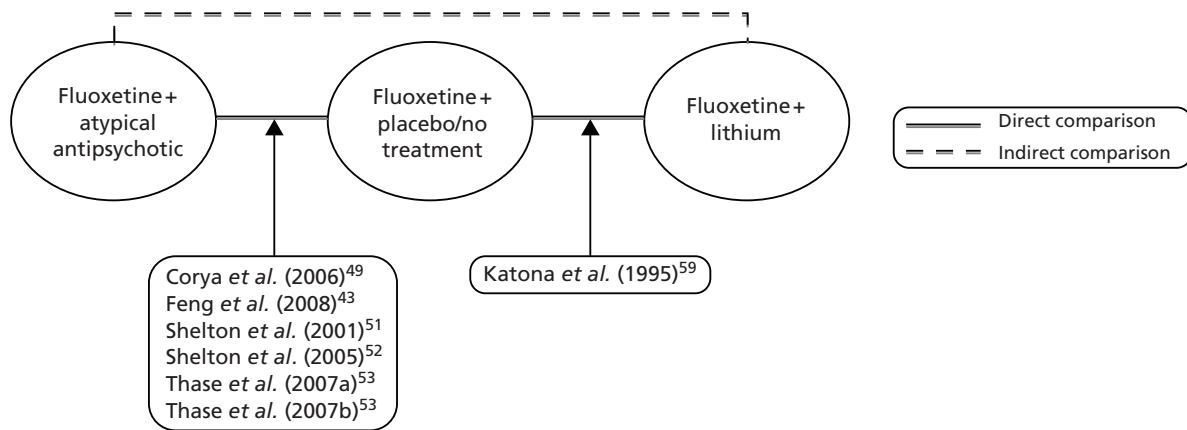


FIGURE 10 Network of RCTs included in the primary network meta-analysis.

In addition, the three RCTs^{47,48,50} reporting antidepressant + aripiprazole compared with antidepressant alone, the RCT⁴⁶ reporting antidepressant + quetiapine compared with antidepressant alone, and the RCT³⁰ comparing antidepressant + quetiapine with antidepressant + lithium were included in a class-based sensitivity analysis.

The results from the MTC for the analyses of SSRI + AAP compared with SSRI alone, and SSRI + lithium compared with SSRI alone are reported in *Table 5*, together with the data for the main comparison of interest: SSRI + AAP compared with SSRI + lithium. A summary table detailing which trials were included in each analysis for each of the outcomes is presented in *Appendix 6*.

The consistency of the evidence was assessed for each outcome using the posterior mean residual deviance, which should approximate the number of unconstrained data points in a good-fitting model. For all outcomes other than mean change in MADRS and the sensitivity analysis for discontinuations, the numbers of unconstrained data points were similar to the value of the posterior mean residual deviance, suggesting that the chosen models were good-fitting. The models used for mean change in MADRS and the sensitivity analysis for discontinuations will be discussed further with the results below. Where a random-effects model was deemed the best fit, the degree of heterogeneity was investigated by evaluating the posterior mean tau-squared (hereafter referred to as 'tau').³⁹ The values of tau varied between 0.16 (sensitivity analysis 1) and 0.47 (mean change in MADRS) for all but two outcomes (dry mouth and headache), suggesting the presence of some heterogeneity.⁶³ The values of tau for dry mouth and headache were higher (0.62 and 0.55, respectively), suggesting moderate heterogeneity in these analyses, although there were data from only four trials in each analysis.⁶³

Response

Seven trials^{43,49,51–53,59} were suitable for inclusion in this analysis. A decision was taken to conduct two separate analyses for the outcome of response owing to the trial informing the comparison with lithium⁵⁹ reporting response using two different criteria. Neither of these criteria was directly comparable with the definition of response used in the other trials included in this analysis. A standard definition of response used in RCTs for depression is a $\geq 50\%$ reduction in MADRS or HAMD score compared with baseline score.⁶⁴ The definitions used in the Katona *et al.* trial⁵⁹ are described in *Box 4*.

For the purposes of comparability with the other trials in this analysis, it is the response from Phase II baseline that is critical for the Katona *et al.* trial.⁵⁹ It is difficult to fully assess the impact that the differences among the trial definitions of response may have on the overall results for this outcome, but this will be discussed and explored further in the discussion section (see *Assessment of effectiveness*, below).

TABLE 5 Summary of results from the network meta-analysis

Outcome	SSRI + AAP vs. SSRI alone: OR (95% CrI)	SSRI + lithium vs. SSRI alone: OR (95% CrI)	SSRI + lithium vs. SSRI + AAP: OR (95% CrI)
Response (using lithium a priori data) ^b	1.96 ^a (1.01 to 3.88)	2.26 (0.24 to 8.96)	1.29 (0.11 to 5.32)
Response (using lithium post hoc data) ^{b,c}	1.99 ^a (1.03 to 4.01)	7.44 (0.59 to 35.30)	4.15 (0.25 to 20.34)
Mean change in MADRS score from baseline ^{b,d}	-2.19 ^a (-3.62 to -0.79)	-3.11 (-10.67 to 5.13)	-1.47 (-9.10 to 6.41)
Withdrawals (all cause) ^e	1.27 (0.90 to 1.75)	0.92 (0.13 to 3.32)	0.74 (0.10 to 2.66)
Total adverse events ^e	1.71 (0.93 to 2.94)	2.87 (0.45 to 10.66)	1.84 (0.24 to 7.05)
Somnolence ^b	2.75 (0.75 to 6.87)	1.14 (0.01 to 7.29)	0.58 (0.004 to 4.01)
Dry mouth ^b	1.95 (0.45 to 5.51)	53.24 (0.13 to 124.8)	47.47 (0.06 to 92.44)
Headache ^b	0.32 ^a (0.08 to 0.86)	4.48 (0.14 to 27.37)	20.49 (0.41 to 134.6)
Sensitivity analysis 1 ^b	1.99 ^a (1.41 to 2.93)	2.11 (0.86 to 4.95)	1.07 (0.44 to 2.32)
Sensitivity analysis 2 ^b	3.13 (0.85 to 10.29)	7.96 (0.56 to 37.70)	3.73 (0.13 to 19.01)
Sensitivity analysis 3 ^b	1.64 (0.89 to 3.32)	7.60 (0.68 to 34.57)	5.18 (0.36 to 23.24)
Sensitivity analysis 4 ^{b,d}	-2.82 ^a (-3.76 to -1.92)	-1.89 (-5.36 to 1.38)	1.27 (-1.88 to 4.68)
Sensitivity analysis 5 ^e	1.24 (0.97 to 1.56)	0.91 (0.12 to 3.17)	0.75 (0.10 to 2.62)

CrI, credible interval.

a Statistically significant at $p < 0.05$.

b Random-effects model.

c Primary analysis for this review.

d MD.

e Fixed-effects model.

Notes

Sensitivity analysis 1: Class-based sensitivity analysis for response.

Sensitivity analysis 2: Sensitivity analysis for response limiting to trials reporting failure to two or more antidepressants in the current episode of depression.

Sensitivity analysis 3: Sensitivity analysis for response limiting to trials reporting response based on the MADRS scale.

Sensitivity analysis 4: Class-based sensitivity analysis for mean change in MADRS.

Sensitivity analysis 5: Class-based sensitivity analysis for discontinuations.

The results of the MTC for the outcome of response using the a priori data from Katona *et al.*⁵⁹ for the comparison of fluoxetine + lithium with fluoxetine + olanzapine were a non-significant trend in favour of treatment with lithium [OR 1.29; 95% credible interval (CrI) 0.11 to 5.32], although this was not statistically significant. When the lithium data that met the post hoc definition of response from Katona *et al.*⁵⁹ were used (the primary analysis for this review) there was a larger treatment effect favouring lithium augmentation over AAP augmentation but the difference between groups remained non-significant (OR 4.15; 95% CrI 0.25 to 20.34). In both instances, the best-fitting model, as determined by the model with the lowest DIC, was a random rather than a fixed-effects model (75 vs. 77 for lithium a priori analysis and 74 vs. 77 for lithium post hoc analysis).

Mean change in MADRS scores

Five RCTs provided data for this analysis [Corya *et al.* (2006),⁴⁹ Katona *et al.* (1995),⁵⁹ Shelton *et al.* (2005)⁵² and Thase *et al.* (2007a+b)⁵³] and the preferred model was a random rather than a fixed-effects model (DIC 26 vs. 28, respectively). The MTC resulted in a MD of -1.47 (95% CrI -9.10 to 6.41) for the mean change in MADRS score from baseline for fluoxetine + lithium compared with fluoxetine + olanzapine,

which suggests a statistically non-significant trend in favour of lithium augmentation. However, the wide 95% CrI indicates a large amount of uncertainty in this estimate of treatment effect and so the results should be interpreted with caution.

Remission

There were no suitable data to conduct an analysis for this outcome.

Quality of life

There were no suitable data to conduct an analysis for this outcome.

Withdrawals (all cause)

Six RCTs^{49,51–53,59} were suitable for inclusion in this analysis and the preferred model was a fixed-effects model rather than a random-effects model (DIC 58 vs. 60, respectively). The results of the MTC for the comparison of fluoxetine + lithium with fluoxetine + olanzapine suggest a statistically non-significant trend in favour of treatment with SSRI + lithium, i.e. withdrawals are less likely (OR 0.74; 95% CrI 0.10 to 2.66).

Withdrawals due to adverse events

There were insufficient data to conduct an analysis for this outcome.

Adverse events

Three RCTs^{43,52,59} provided data for the analysis of the total number of people reporting adverse events. The fixed- rather than the random-effects model was best-fitting (DIC 34.0 vs. 34.3, respectively) and resulted in a statistically non-significant benefit for fluoxetine + lithium compared with fluoxetine + olanzapine (OR 1.84; 95% CrI 0.24 to 7.05). However, the number of events informing this analysis is small and so the results should be interpreted with caution.

The individual adverse events with sufficient data for comparison in the MTC were limited to somnolence, dry mouth and headache. All three outcomes were analysed using a random-effects model and they all included data from four trials.^{49,53,59} The ORs for the comparison of fluoxetine + lithium with fluoxetine + olanzapine were as follows:

- *somnolence* OR 0.58 (95% CrI 0.004 to 4.01)
- *dry mouth* OR 47.47 (95% CrI 0.06 to 92.44)
- *headache* OR 20.49 (95% CrI 0.41 to 134.6).

The 95% CrIs for all three of these outcomes are extremely wide and thus the estimates for the ORs are extremely uncertain and should be interpreted with caution.

Relapse rate

There were no suitable data to conduct an analysis for this outcome.

Mortality

There were no suitable data to conduct an analysis for this outcome.

A priori subgroup analyses

The following prespecified subgroup analyses could not be conducted owing to an absence of suitable data from the clinical effectiveness trials included in this review:

- (a) different durations of depression (i.e. time since first onset of current episode of depression)
- (b) classes of previous antidepressants (e.g. SSRI or tricyclic antidepressant)
- (c) sex
- (d) age (i.e. those <75 years and those \geq 75 years)
- (e) people with different severities of depression.

Sensitivity analyses

Details of the five sensitivity analyses are as follows.

Sensitivity analysis 1 This sensitivity analysis assumed that there is a class effect for all SSRIs and also a class effect for all AAPs. It enabled the inclusion of all 12 RCTs identified as suitable for inclusion in this review.^{30,43,46–53,59} The random rather than the fixed-effects model was best fitting (DIC 141 vs. 143, respectively) and suggests a statistically non-significant trend favouring treatment with SSRI+lithium compared with SSRI+AAP for the outcome of response (OR 1.07; 95% CrI 0.44 to 2.32), based on \geq 50% reduction in MADRS or HAMD score and the post hoc definition of response used in the trial by Katona *et al.*⁵⁹ This is consistent with the results from the primary analysis.

Sensitivity analysis 2 This sensitivity analysis was decided post hoc to assess the impact of including some trials with patients with a failure to respond to two or more antidepressants of which only one treatment failure was in the current episode of depression (the remaining failures were in historical episodes of depression). The trials included in the sensitivity analysis were limited to those in which all patients had a failure to respond to two or more antidepressants in their current episode of depression. Four trials^{51,53,59} met this strict criterion. A random rather than fixed-effects model was best fitting for this outcome (DIC 35 vs. 36, respectively). The results favour augmentation of SSRI with lithium over augmentation with olanzapine (OR 3.73; 95% CrI 0.13 to 19.01), although this result was non-significant. This is consistent with the results from the primary analysis.

Sensitivity analysis 3 This sensitivity analysis was conducted to assess the impact of limiting the primary analysis to trials reporting response based on the MADRS scale. It resulted in the exclusion of only one previously included RCT,⁴³ leaving a total of six trials^{49,51–53,59} included in this sensitivity analysis. The random rather than fixed-effects model was best fitting (DIC 64.1 vs. 64.5, respectively) and resulted in a statistically non-significant trend in favour of treatment with SSRI+lithium compared with SSRI+olanzapine (OR 5.18; 95% CI 0.36 to 23.24). This is consistent with the results of the primary analysis.

Sensitivity analysis 4 This analysis assumed a class effect for all SSRIs and a class effect for all AAPs and was used to assess the impact of the decision to use a non-class-based assumption for the primary analysis on the outcome measure of mean change in MADRS score. This sensitivity analysis was conducted in addition to the sensitivity analyses on the primary outcome measure to enable a class-based sensitivity analysis to be performed in the economic model.

Eight RCTs^{30,47–50,52,53,59} were suitable for inclusion in this sensitivity analysis. These trials comprised data from the six trials^{30,49,52,53,59} in the primary analysis on the comparison of SSRI (or venlafaxine) + quetiapine with SSRI (or venlafaxine) + lithium, as well as data from the three SSRI (or venlafaxine) + aripiprazole trials compared with SSRI (or venlafaxine)-alone trials.^{47,48,50}

The data from two of the SSRI + aripiprazole compared with SSRI-alone trials,^{47,50} were available as a pooled analysis⁶² that was broken down by individual SSRI, and so these data were included as this enabled the exclusion of the data from patients with baseline venlafaxine treatment, thus reducing the amount of potential clinical heterogeneity. The random rather than fixed-effects model was deemed to be the best fitting (DIC 45 vs. 47, respectively) and resulted in a MD of 1.27 (95% CrI –1.88 to 4.68) for SSRI+lithium compared with SSRI+AAP. This suggests a statistically non-significant trend in favour of treatment with SSRI+AAP, which is in contrast to the results of the primary analysis, in which the trend was in favour of

treatment with SSRI + lithium. However, both analyses demonstrate no significant difference in treatment effect between treatment with SSRI + lithium and with SSRI + AAP.

Sensitivity analysis 5 The sensitivity analysis assumed a class effect for all SSRIs and a class effect for all AAPs for the outcome measure of withdrawals (all cause). This sensitivity analysis was conducted to provide the additional clinical data required to enable a class-based sensitivity analysis to be performed in the economic model.

Ten trials were suitable for inclusion in the analysis.^{46–53,59} These trials included the six trials^{49,51–53,59} from the primary analysis, the three trials^{47,48,50} with data on SSRI (or venlafaxine) + aripiprazole compared with SSRI (or venlafaxine) alone and the Mattingly *et al.*⁴⁶ trial with data on SSRI (or SNRI) + quetiapine compared with SSRI (or SNRI) alone.

The trial³⁰ for the comparison of SSRI (or venlafaxine) + quetiapine with SSRI (or venlafaxine) + lithium did not report subgroup data for people with TRD – defined as failure to respond to two or more antidepressants in the current episode of depression – and so could not be included in this analysis.

The fixed rather than random-effects model was the best fitting (DIC 94 vs. 96, respectively). The resulting OR for the comparison of SSRI + lithium with SSRI + AAP was 0.75 (95% CrI 0.10 to 2.62). This suggests a statistically non-significant trend in favour of treatment with SSRI + lithium (i.e. withdrawals are less likely). This result is consistent with the results of the primary analysis.

Discussion and summary findings

The available clinical effectiveness data informing the comparison of SSRI + AAP with SSRI + lithium in the primary analysis were based on fluoxetine + olanzapine^{43,49,51–53} compared with fluoxetine + lithium.⁵⁹ The results from the MTC of the star-shaped network demonstrate no significant differences between treatment regimens for any of the outcomes assessed. A non-significant trend in benefit was observed for the lithium-based augmentation strategy compared with the olanzapine-based augmentation strategy for response, mean change in MADRS from baseline, and fewer discontinuations. The results of the MTC also demonstrated a non-significant trend in favour of the olanzapine-based augmentation strategy compared with the lithium-based augmentation strategy for fewer adverse events. However, care should be taken when interpreting non-significant results.

When the results of the MTC are compared with the individual results for the pairwise meta-analyses there is general agreement with the results obtained when SSRI is used as the baseline. This suggests a reasonable fit of the modelling approach used within the MTC. The radiating star shape of the network means that only the trials providing the results from the pairwise meta-analyses are providing the results within the MTC. The results for the lithium-based augmentation strategy compared with SSRI alone in the MTC tend to have wider 95% CrIs than the 95% CIs provided from the single trial informing this comparison.⁵⁹ This is probably owing to the random-effects model tending to be the preferred model for the MTC outcomes assessed.

The results for both the pairwise meta-analysis and MTC estimates of SSRI + AAP compared with SSRI alone showed a statistically significant benefit in favour of augmentation with AAP for the outcomes of response and mean change in MADRS score. The equivalent results for SSRI + lithium compared with SSRI alone showed a statistically non-significant trend in favour of augmentation with lithium. The results for lithium augmentation could be considered inconclusive because they do not reach statistical significance, although it should be noted that they are based on data from only a small subgroup of patients in one RCT.⁵⁹ In addition, other publications have reported results that suggest lithium is an effective augmentation strategy.^{60,65,66}

A recent meta-analysis by Crossley *et al.*⁶⁰ included 10 RCTs and demonstrated a statistically significant benefit in terms of response rate with lithium augmentation compared with placebo (OR 3.11; 95% CI 1.80 to 5.37). This meta-analysis is not entirely comparable with the patient population under review here because it included patients with bipolar disorder, patients on various antidepressants (including tricyclic antidepressants), and patients with a minimum of one previous antidepressant failure. However, the Crossley *et al.*⁶⁰ meta-analysis does provide evidence to suggest that lithium is an effective augmentation agent in TRD and thus is supportive of the results from the MTC. In addition, the results from the full trial population for the Katona *et al.*⁵⁹ RCT included in the MTC demonstrate a statistically significant benefit in terms of response with lithium augmentation compared with placebo (OR 3.21; 95% CI 1.09 to 9.48).

Relapse rates and mortality were prespecified as outcomes of interest for this review. However, none of the identified trials reported comparable mortality or relapse rate data. In addition, extremely limited subgroup data, if any, were reported for the trials included in the clinical effectiveness review and no trial reported suitable subgroup data for the prespecified subgroup analyses.

The trials included in this review were validated against the trials included in CG90¹⁶ and those included in two separate systematic reviews; one for the comparison of AAP augmentation with placebo in MDD¹¹ and the other for lithium augmentation compared with placebo in TRD.⁶⁰ All three publications included additional trials compared with this review, but none of the additional trials was found to be suitable for inclusion. Moreover, each trial was excluded for multiple reasons based on the inclusion criteria for this review. Mostly, the additional RCTs did not meet the following criteria: the augmentation of SSRI as baseline therapy;^{67–82} a 4-week minimum duration of treatment;^{74–80,81} or two or more failures of antidepressant therapy in the current episode of depression^{67–73} (it is unclear how many trials in Crossley *et al.*⁶⁰ are affected by this criterion). The results of both the Crossley *et al.*⁶⁰ meta-analysis evaluating lithium compared with placebo augmentation and the Nelson *et al.*¹¹ review of AAP compared with placebo augmentation demonstrated that the respective augmentation agents were statistically significantly more effective than placebo at achieving treatment response in people with MDD or TRD.

Sensitivity analyses

The sensitivity analysis based on a class effect (sensitivity analysis 1) allowed the inclusion of two additional AAPs (aripiprazole and quetiapine) in the MTC. For the outcomes assessed (response), the non-significant trend favouring lithium augmentation observed in the primary analysis is diminished (mean OR changes from 4.15 to 1.07) and remains non-significant. This could be supportive evidence for no clinically meaningful difference between the two augmentation strategies, a reflection of the lack of information available for the comparison, or an indication that the assumption of a class effect is flawed and that there is a difference between the individual treatments within a class. With regards to the last concern, the trial providing information on quetiapine in the MTC³⁰ demonstrated a non-significant trend in favour of a quetiapine-based augmentation strategy compared with lithium-based augmentation (OR 1.25; 95% CI 0.74 to 2.12). The result of this trial is thus in contrast to the result of the MTC and therefore suggests a difference in treatment effect between the different AAPs.

In the primary analysis, RCTs were included if the majority of patients in each individual trial had experienced two or more previous failures on an antidepressant in their current episode. A sensitivity analysis (sensitivity analysis 2) was conducted to determine if restricting this criterion to RCTs in which the whole population had experienced two or more previous failures in their current episode had a substantial impact on the results of the MTC. The results were consistent with the primary analysis for response, that is, a non-significant trend in favour of lithium-based augmentation (using lithium post hoc data).

A further sensitivity analysis (sensitivity analysis 3) was conducted as an exploratory analysis to assess the impact of limiting trials in the primary analysis to those reporting response on the MADRS scale. It resulted in the exclusion of a single RCT⁴³ that was conducted in China and in which a different classification

system had been used for the diagnosis of depression. The other trials in the analysis were mainly based in the USA and used the DSM classification for depression. However, the results of this sensitivity analysis were in keeping with the primary analysis for response.

An additional sensitivity analysis (sensitivity analysis 4) was conducted for the economic model on the outcome of mean change in MADRS score from baseline assuming a class effect. Although the results were similar to the primary analysis in terms of no significant difference, the trend changed from being in favour of lithium-based augmentation to being in favour of AAP-based augmentation. This change in trend is associated with the inclusion of the Franco *et al.* trial.³⁰ As discussed earlier, this trial demonstrates a non-significant benefit for response for quetiapine-based augmentation over lithium-based augmentation and the same trend is observed for reduction in MADRS scores. However, a MD in MADRS of – 1.47 in favour of lithium-based augmentation from the primary analysis or a MD of 1.16 in MADRS in favour of AAP-based augmentation is unlikely to be clinically meaningful.⁸³

The final sensitivity analysis (sensitivity analysis 5) was also conducted for the benefit of the economic model. This evaluated the impact of a class-based assessment of withdrawals (all cause). The results were similar to the primary analysis and demonstrated no significant difference in the risk of patients withdrawing from treatment with either augmentation strategy.

Limitations

The major limitation in this review of the clinical effectiveness of SSRI + AAP compared with SSRI + lithium in patients with TRD after failing on two or more previous antidepressants was the absence of a direct comparison in a RCT. The number of RCTs available with data in the specified population, and at the right stage in their disease management, for comparison of the two augmentation strategies within a MTC was small. In the primary analysis this consisted of six RCTs for fluoxetine + olanzapine^{43,49,51–53} and one RCT for fluoxetine + lithium.⁵⁹ In addition, this comparison was made possible only by including a surrogate trial in a slightly less-severe population to allow the analysis to contain the lithium trial.⁵⁹ This trial comprised around 50% of people who did not meet the definition of TRD for this review, as they had failed only one prior antidepressant in their current episode of depression.⁵⁹ Patients in this trial could be less severe than patients included for the fluoxetine + olanzapine comparison and so had a greater potential to respond to treatment. However, as this potential to have a greater response would be similar in both treatment groups within the trial the impact of this potential source of bias on the analysis is difficult to determine.

Similarly, the RCTs identified for the primary analysis provided sufficient data for only a 'stepwise' indirect comparison of the data, i.e. A versus B versus C, without the possibility of drawing further strength and cohesion in the primary analysis that a trial of A versus C would have provided. This is compounded by the SSRI + lithium compared with SSRI alone link being formed by only a single, small trial ($n = 33$).⁵⁹ The impact of this can be observed in the relatively high levels of uncertainty obtained in any of the results of the MTC.

In addition, the single trial informing the comparison of SSRI + lithium with SSRI alone has what may be considered counterintuitive results for withdrawals (all cause), as fewer withdrawals occur in this study with fluoxetine + lithium than fluoxetine alone (4/17 vs. 5/16, respectively). Lithium is known to have a different side effect profile to the SSRIs, with the additional inconvenience of requiring frequent blood tests during treatment.¹⁹ The results for this outcome should be treated with caution.

None of the trials identified for inclusion in the review provided information on all the outcomes assessed, which is another source of uncertainty around the estimated treatment effects. However, the fundamental issue with regards to uncertainty is most profoundly driven by the single small trial identified for the lithium augmentation strategy.

Definitions of response to treatment were also different among the trials identified for inclusion. The definition of response used in this review was $\geq 50\%$ reduction in MADRS/HAMD score from baseline. This single trial informing the lithium augmentation strategy described two different definitions used (one a priori and one post hoc),⁵⁹ neither of which was consistent with the one used in this review. This potentially makes the results from this outcome less reliable for the comparison of the two augmentation strategies.

Finally, a further limitation is that the network of RCTs identified by the systematic review was limited to those treatments within the scope of the review. Important 'linking' studies (e.g. tricyclic antidepressants or SNRIs used as a baseline for augmentation) could have provided additional information to enhance the precision and reliability of the estimates generated by the MTC.

Overall

The results of this review support the conclusion that augmentation of SSRIs with lithium or an AAP is likely to be beneficial in people with TRD, defined as failure to respond to two or more antidepressants in their current episode of depression. However, based on the limited number of RCTs identified in this research, the clinical evaluation suggests there is no statistically significant difference between the two augmentation strategies. There is a general paucity of RCT data available in patients with TRD for SSRI + lithium and SSRI + AAP.

Chapter 4 Assessment of cost-effectiveness

In addition to addressing the decision problem that is the focus of this review, this report aims to compare the cost-effectiveness of the augmentation of SSRI therapy with lithium with augmentation with an AAP in a patient population with TRD (the decision problem is discussed further in *Chapter 2, Decision problem*). Assessment of the comparative cost-effectiveness of these two augmentation strategies is required to facilitate the effective and efficient allocation of health-care resources within the NHS.

It was anticipated that the evidence relating to the cost-effectiveness of augmenting SSRI treatment with lithium or with an AAP in a TRD patient population would be limited. Therefore, a de novo economic evaluation was carried out to compare the expected costs and benefits of the two augmentation strategies of interest. To ensure all of the relevant evidence from the economic literature was incorporated into the model, two systematic literature reviews were conducted alongside model development. These reviews aimed to identify:

- economic evaluations that could inform the methodological approach to the development of a de novo model or improve understanding of the economic consequences of the disease area (economic literature review)
- utility values associated with depression and treatments for depression (QoL literature review).

The following sections in this chapter and the first six sections in *Chapter 5* (see *Introduction, Model overview, Model structure, Effectiveness data, Costs, and Sensitivity analysis*) describe the methodology and results of each review and detail the structure, assumptions and results of the de novo economic evaluation.

Economic literature review

A systematic review of the literature was carried out to identify potentially relevant economic evaluations in the management of TRD. As discussed above, it was anticipated that the cost-effectiveness evidence relating to augmentation strategies in TRD would be limited. Consequently, the decision was taken to broaden the search to economic evaluations of any intervention in patients with TRD. Therefore, the aim of the economic literature review was to identify evidence that could:

- inform the methodological approach to the development of a de novo decision-analytic model or
- improve understanding of the economic consequences of the disease area.

Literature search terms and strategies used to identify cost-effectiveness studies relevant to the decision problem were based on validated search strategies developed by Haynes *et al.*⁸⁴ and Dickersin *et al.*⁸⁵ Multiple databases encompassing medical and economic literature were searched to maximise the potential of capturing relevant studies. Databases searched were:

- EMBASE (for the period 1988 to August 2011)
- MEDLINE, including MEDLINE In-Process & Other Non-Indexed Citations (for the period 1950 to August 2011)
- Cochrane Central Register of Controlled Trials (CENTRAL) (from inception to August 2011)
- HTA database (from inception to August 2011)
- NHS Economic Evaluation Database (NHS EED) accessed via Wiley Online Library (for the period 1999 to August 2011)
- PsycINFO (from inception to August 2011).

Details of the search strategy are provided in *Appendix 7*. The search terms used covered condition, population and intervention and no country or language restrictions were applied. All references were exported to the Reference Manager (Thomson ResearchSoft, San Francisco, CA, USA) bibliographic database and deduplicated. In addition, the reference lists of identified systematic reviews were hand-searched for references. Inclusion and exclusion criteria, specific to the cost-effectiveness literature review, were developed and are displayed in *Table 6*.

The search identified 438 studies, of which 104 were duplicates. One reviewer (LN) carried out the first appraisal of the title and abstract of the 334 studies (level one screening) for potential inclusion. At this stage, 301 papers were excluded. The 33 abstracts identified as being potentially relevant were independently appraised (level two screening) by a second reviewer (NT) and 10 abstracts were excluded at this stage. The full-text publication of 23 studies identified as potentially relevant by both reviewers (LN and NT) were ordered and independently assessed for inclusion, using the criteria outlined in *Table 6*. Following the full-text review, 19 studies were excluded and four economic evaluations⁸⁶⁻⁸⁹ were identified for final inclusion (*Table 7*). The study selection process is summarised in *Figure 11*.

Overview of included economic evaluations

The four economic evaluations identified in the systematic literature review were carried out in various countries: Scotland,⁸⁶ Singapore,⁸⁷ Thailand⁸⁸ and the USA.⁸⁹ The study by Benedict *et al.*⁸⁶ was from the perspective of the Scottish NHS. Therefore, the costs and resources considered in the evaluation by Benedict *et al.*⁸⁶ were the most relevant to the decision problem that is the focus of this review. However, it is well established that TRD is associated with a substantial societal burden.^{16,22} Benedict *et al.*⁸⁶ considered the impact of the societal perspective in sensitivity analysis. The US study by Simpson *et al.*⁸⁹ took a similar approach to that of Benedict *et al.*⁸⁶ (payer perspective as the base case and societal perspective in sensitivity analysis), whereas the studies carried out in Thailand⁸⁸ and Singapore⁸⁷ were conducted from a payer and societal perspective, respectively. A summary of the four included economic evaluations is given in *Table 7*; the quality of each evaluation (individually discussed below; see *Narrative review of included studies*) was assessed using the Philips checklist⁹⁰ (the checklist for each study can be found in *Appendix 8*).

All of the retrieved studies were recent, being published in 2009 or 2010, and are therefore likely to reflect current clinical practice in relation to TRD. Each study included a population of patients who were treatment resistant (either the whole study population or a subgroup); TRD is defined as per the population criteria for this review.

The health states used in each of the identified studies informed the number and type of health states included in the de novo model (see *Chapter 5, Model overview and Model structure*). However, the

TABLE 6 Inclusion and exclusion criteria used in the review of the economic evaluation literature

Criteria	Inclusion	Exclusion
Type of study	Full economic evaluation	Systematic review
Population	People with TRD, defined in line with the criteria for this review (patients who have failed to sufficiently respond to two or more adequate lines of SSRI therapy in their current episode of depression)	People with psychotic conditions, including bipolar I or II disorder
Geographical location	Publications from any country	None
Interventions	Any	None
Outcomes of interest	QALYs, other health outcome measures and expected costs	None

QALY, quality-adjusted life-year.

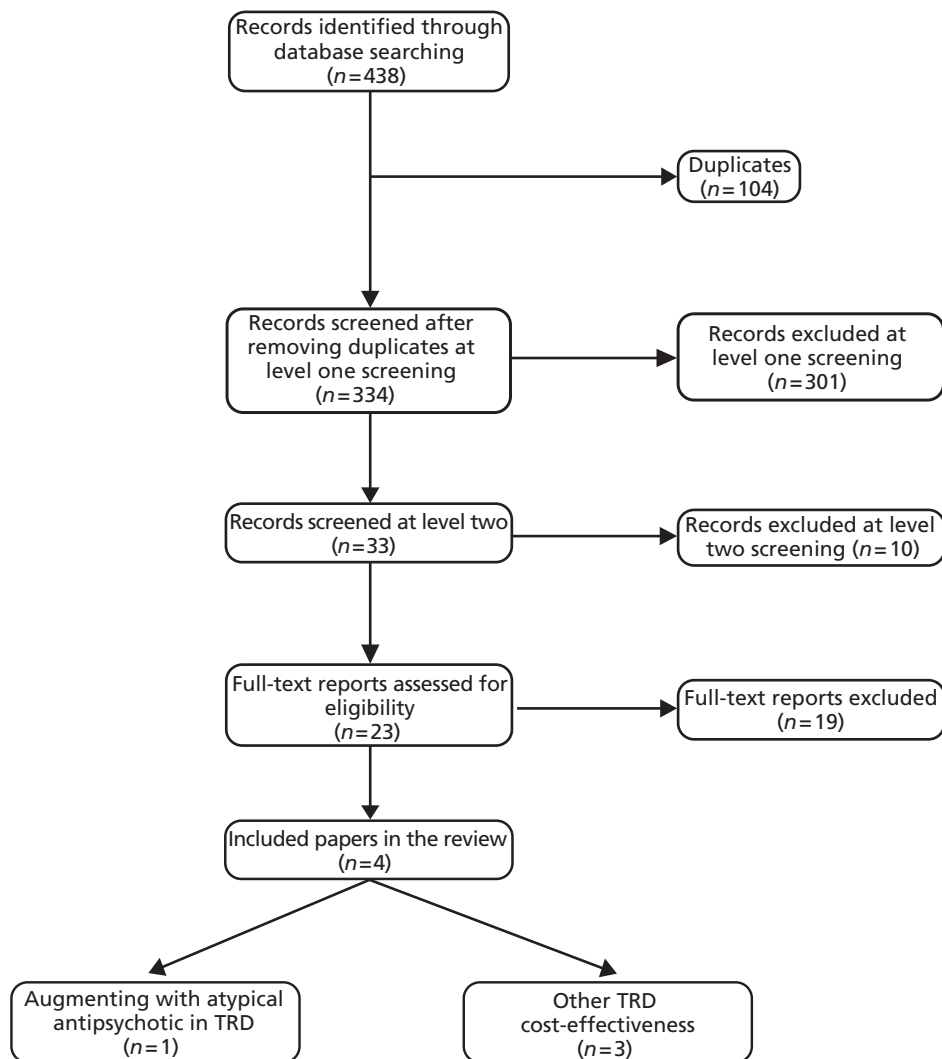


FIGURE 11 Number of cost-effectiveness studies identified and excluded at each stage in the appraisal process.

methodological approach to the simulation of patients varied across the identified studies. Two of the studies used a decision tree model,^{87,88} one study used a Markov model⁸⁶ and the last⁸⁹ used a hybrid model, combining a decision tree to simulate the acute phase of treatment and a Markov component to simulate the maintenance phase of treatment. The hybrid approach used by Simpson *et al.*⁸⁹ accounted for both the short (duration of acute therapy) and medium (1-year time horizon) term implications of treatment. Therefore, the study by Simpson *et al.*⁸⁹ was considered to be the most relevant to the current decision problem. Furthermore, none of the identified studies considered the longer-term (> 1 year after acute treatment) implications of treatment and the time horizons evaluated ranged from 6 weeks⁸⁸ to 1 year.^{86,89} This reflects the paucity of long-term follow-up data available for TRD patients.¹⁶

Only one study⁸⁸ evaluated augmentation of antidepressant therapy [augmented with an AAP (aripiprazole)]; the three remaining studies^{86,87,89} evaluated antidepressant monotherapy regimens compared with either each other or transcranial magnetic stimulation (TMS; Neuronetics, Inc., Malvern, PA, USA). Incremental cost per quality-adjusted life-year (QALY) gained was assessed in all studies except that by Xie *et al.*,⁸⁷ which limited the outcomes assessed to the percentage of primary and secondary care patients achieving remission and the associated costs.

TABLE 7 Summary of included economic evaluations

Author, year, country	Perspective	Model type	Patient population	Intervention vs. comparator	Outcomes	ICER (per QALY gained)
Leelahanjai <i>et al.</i> (2010), ⁸⁸ Thailand	Payer	Decision tree	Patients with non-psychotic MDD who have incompletely responded to at least two antidepressants, including one prospective 8-week antidepressant treatment course	Augmentation with aripiprazole vs. augmentation with placebo	Incremental cost per remission and incremental cost per QALY	3201 Thai baht (2009)
Xie <i>et al.</i> (2009), ⁸⁷ Singapore	Societal	Decision tree	Patients with MADRS score of ≥ 22 treated either in primary care or secondary care	Escitalopram vs. venlafaxine or fluvoxamine	Percentage of primary and secondary care patients achieving remission and associated costs	N/A
Benedict <i>et al.</i> (2010), ⁸⁶ Scotland	Payer and societal	Markov	Patients with moderate to severe MDD (HAMID-17 score of ≥ 19) and patients with severe MDD (HAMID-17 score of ≥ 25) were considered separately	Duloxetine (Cymbalta, Eli Lilly) vs. venlafaxine ER or mirtazapine or SSRI	Incremental cost per QALY	<p>Primary care population (HAMID-17 score of ≥ 19):</p> <ul style="list-style-type: none"> Duloxetine vs. venlafaxine ER: duloxetine dominant Duloxetine vs. mirtazapine: £2353 Duloxetine vs. SSRIs: £6304 <p>Secondary care population (HAMID-17 of ≥ 25):</p> <ul style="list-style-type: none"> Duloxetine vs. venlafaxine ER: duloxetine dominant Duloxetine vs. mirtazapine: duloxetine dominant
Simpson <i>et al.</i> (2009), ⁸⁹ USA	Payer and societal	Hybrid (decision tree plus Markov)	Patients with pharmacotherapy-resistant major depression in the current illness episode as measured by the ATHF (moderate to severe depression and moderate to severe resistance)	TMS vs. sham TMS TMS vs. pharmacotherapy as usual	Incremental cost per QALY	<p>Direct costs only:</p> <ul style="list-style-type: none"> TMS vs. sham treatment US\$34,999 <p>Including productivity costs:</p> <ul style="list-style-type: none"> TMS vs. sham treatment US\$6667 TMS dominated pharmacotherapy

ATHF, antidepressant treatment history form; ER, extended release; ICER, incremental cost-effectiveness ratio; N/A, not applicable.

Narrative review of included studies

Leelahanj *et al.* 2010

Leelahanj *et al.*⁸⁸ undertook a cost-effectiveness analysis comparing aripiprazole with placebo as adjunctive therapy to antidepressants in the treatment of non-psychotic MDD. The study included data from two trials in patients with MDD and with a history of inadequate response to between one and three adequate (appropriate dose given for > 6 weeks) antidepressant trials. Patients who then subsequently failed (achieved partial response) to respond to treatment in a prospective 8-week trial of a different antidepressant (to those previously administered) were randomised to adjunctive therapy with aripiprazole or with placebo. A cost-effectiveness analysis was undertaken from a Thai payer perspective, beginning at the point of patient randomisation.

Model structure and assumptions

A decision tree model was constructed to calculate the comparative costs and effects of adjunctive therapy with aripiprazole or placebo in Thailand. The 6-week time horizon of the cost-effectiveness analysis matched the duration of the randomised phase of the trials from which data were taken for analysis. Patients entered the model at the point of randomisation and received either aripiprazole or placebo. After randomisation, patients could remit, discontinue treatment, or remain on therapy without achieving remission. Patients who did not remit were assumed to be hospitalised for ECT, after which all patients were assumed to be in a state of remission, including those who discontinued their therapy. The primary outcome of the model was cost per remission, with remission defined as a total MADRS score¹⁴ of ≤ 10 . Cost per QALY was assessed as a secondary model outcome.

Efficacy data

In the model, the efficacy of aripiprazole is parameterised solely with data on remission and discontinuation. Remission and discontinuation rates are from a pooled analysis of two identical US-based RCTs of adjunctive aripiprazole in patients with MDD with inadequate response to antidepressant therapy. The authors reported that a utility value of 0.8 was applied to the state of remission (the only end health state), from which the overall cost per QALY is then calculated. This utility value is from a study by Revicki *et al.*⁹¹

Resource use and cost data

The analysis was conducted from a health-care payer perspective and therefore accounts for only direct health-care costs, including drug costs (aripiprazole and antidepressants), hospitalisation costs, and the cost of ECT. Patients achieving remission after augmentation therapy incurred the cost of 6 weeks of antidepressant therapy plus the cost of their augmentation drug (aripiprazole = 195.5 baht per day, placebo = 0 baht per day). Patients who did not remit or discontinue from augmentation therapy incurred 6 weeks of augmentation and antidepressant therapy, plus a further 20 days of antidepressant therapy. Patients discontinuing augmentation therapy incurred only 20 days of antidepressant therapy. In addition, patients who did not remit after augmentation therapy also incurred the costs of 20 inpatient bed-days and ECT; the estimate of hospitalisation time was obtained from a survey (no further details of the survey are reported) and the model assumed that ECT was performed eight times within the hospitalisation period. All costs were taken from the records of a local hospital.

Summary of results

Remission rates following augmentation and overall costs of augmentation were higher with aripiprazole than with placebo (25.7% and 30,970 baht vs. 15.4% and 28,409 baht, respectively). For aripiprazole compared with placebo, the incremental cost per remission is reported as 2561 baht and the incremental cost per QALY is reported as 3201 baht, respectively. Based on these results, aripiprazole is deemed not to be cost-effective according to the willingness-to-pay threshold in Thailand (not reported). Sensitivity analysis indicated that the remission rate for aripiprazole would need to increase from 25.7% to 34.8% to achieve cost-effectiveness in Thailand. Alternatively, the cost of aripiprazole would need to fall by 48.9%.

Critique

The study by Leelahanaj *et al.*⁸⁸ evaluated the population of interest for this review and used a simple decision tree. However, the study has several flaws, which may limit its applicability to this review. Although the time horizon used matched the length of trial data available, 6 weeks is, as the authors concede, a short time frame and will not inform the likely long- or medium-term cost-effectiveness of aripiprazole. Moreover, the effectiveness data were derived from two US-based studies, which may not be generalisable to a Thai population. Adverse events were omitted from the analysis, which may result in overestimation of the benefit of aripiprazole. Most importantly, the key limitation of this study as a cost-effectiveness analysis is the assumption that all patients remit following treatment with ECT. The consequence of this assumption is that there is no difference in the number of remissions or QALYs gained between treatments. Therefore, this analysis may be considered akin to a cost-minimisation analysis rather than a cost-effectiveness analysis, an approach that is inappropriate between two treatments with significantly different efficacy estimates.

Xie *et al.* 2009

Xie *et al.*⁸⁷ carried out a cost-effectiveness analysis of escitalopram compared with venlafaxine and also with fluvoxamine in the treatment of patients with MDD. The authors constructed a decision-analytic model with two consecutive pathways; the first simulated patients initiating treatment and the second simulated patients who had two failed antidepressant treatments in primary care (who were therefore eligible for augmentation, combination therapy or hospitalisation). All patients had a MADRS score¹⁴ of ≥ 22 and the analysis was conducted from a Singaporean societal perspective. The focus for this review is the second model pathway, which considered patients with a MADRS score¹⁴ of ≥ 22 , treated in secondary care following the failure of two treatments.

Model structure and assumptions

The cost-effectiveness analysis presented in the study considers patients with MDD for 6 months following the initiation of therapy, in accordance with the Singaporean depression clinical guideline 2004.⁹² The model used is a decision tree with two consecutive pathways. The first considers patients as they are initiating therapy in either primary or secondary care, according to clinical practice in Singapore. The second considers patients in secondary care who have failed two initial treatments. The primary outcome of the model was incremental cost per additional remission achieved, with remission defined as a MADRS score¹⁴ of ≤ 12 .

Following initiation, patients could achieve remission, require titration because of lack of efficacy or switch therapy as a result of adverse events. Patients achieving remission had the option of stopping treatment prematurely or continuing with therapy, after which patients either remained in remission or relapsed (with relapses defined as a MADRS score¹⁴ of ≥ 25). The model structure allowed a maximum of one titration and one treatment switch, after which patients who had still not remitted were moved into the secondary care pathway and were eligible for augmentation, combination therapy or hospitalisation. The rates of augmentation, combination therapy and hospitalisation were derived from a survey of GPs and psychiatrists and all outcomes were independent of treatment received. Following augmentation or combination therapy, patients could achieve full/partial response or no response. Responding patients had the option of prematurely ceasing therapy or remaining on therapy, after which they could either remit or relapse.

The final model health states are remission, relapse or hospitalisation.

Efficacy data

Where possible, data from head-to-head trials were used to parameterise the efficacy of escitalopram, venlafaxine and fluvoxamine. However, the absence of a head-to-head trial of escitalopram compared with fluvoxamine led to the use of efficacy data for citalopram as a proxy for the efficacy of fluvoxamine; evidence of non-significant difference in the efficacy and tolerability of citalopram and fluvoxamine was used to justify the appropriateness of this approach.

The rate of remission achieved with each drug is the main effectiveness parameter used in the model. In the comparison of escitalopram with venlafaxine, the remission rates of escitalopram and venlafaxine are

taken from a meta-analysis of patients treated in primary and secondary care.⁹³ However, in the comparison of escitalopram with fluvoxamine, the remission rates associated with escitalopram and fluvoxamine are taken from the head-to-head trial of escitalopram compared with citalopram (with the remission rates of citalopram used as a proxy for the remission rates of fluvoxamine). Following titration, the efficacy of each drug is assumed to decrease slightly, based on evidence from a cited cost-effectiveness analysis.⁹⁴ Remission rates achieved following a treatment switch are assumed to be the same for all therapies, irrespective of the starting therapy.

In the secondary care model pathway, the rate of response (either full or partial) replaces the rate of remission as the main driver of effectiveness. The response rate experienced by patients treated with augmentation or combination therapy is taken from a study of augmentation therapy by Posternak *et al.*⁹⁵ and response rates are assumed to be equal for augmentation or combination treatment strategy (as the outcomes of the secondary care model pathway are independent of treatment allocation).

The rate of relapse is another key clinical parameter in the model that is assumed to be independent of treatment allocation. The rate of relapse demonstrated in studies of citalopram is used to represent the rate of relapse expected in first-line therapy and following augmentation or combination therapy. A slightly lower relapse rate is associated with switch therapy and a higher rate of relapse is associated with patients who prematurely stop antidepressant therapy following remission.

Resource use and cost data

The perspective of this analysis was societal and therefore included both direct and indirect costs, including lost productivity. Estimates of expected resource use were not readily available from the literature and so the authors conducted a survey among GPs and psychiatrists with experience in treating MDD. As part of this survey, the amount of contact with GPs or specialists, hospitalisations and working days lost were estimated for each phase of treatment. The unit cost of drug acquisition, professional consultation (either GP or psychiatrist) and hospitalisation were obtained from local hospitals. The cost of absenteeism was calculated using the human capital method.

Summary of results

Escitalopram dominated both fluvoxamine and venlafaxine, with higher rates of remission and lower costs [escitalopram (68.1%, US\$2845) vs. venlafaxine (66.0%, US\$3176); escitalopram (64.7%, US\$3133) vs. fluvoxamine (60.0%, US\$3297)]. Probabilistic sensitivity analysis (PSA) demonstrated that, out of 5000 runs, 95% resulted in the dominance of escitalopram over venlafaxine and 5% resulted in escitalopram yielding fewer health benefits and lower costs than venlafaxine. Escitalopram dominated fluvoxamine in 98% of runs and resulted in incremental cost-effectiveness ratios (ICERs) of <US\$10,000 in 2% of runs.

No results were presented for the secondary care pathway, as transition probabilities were independent of treatment allocated.

Critique

The decision problem, objective and perspective of this analysis were clearly stated and the outcome of cost per additional remission was appropriate to address the decision problem. A pragmatic decision tree model, originally developed in Europe, was well adapted to apply to the Singaporean health-care system, based on consultation with medical professionals involved in the treatment of MDD. The selected time horizon of 6 months was compatible with the clinical guideline of depression in Singapore. However, this will not have captured the longer-term costs and consequences of the treatment of MDD. The model accounted for patients who were resistant to initial therapies in a pragmatic and consistent manner, based on consultation with the appropriate health-care professionals (HCPs). Generally, the assumptions made regarding the pathway of resistant patients seem to be reasonable. One assumption that contradicts available evidence is the assumption that relapse rates following augmentation or combination therapy would be the same as following first-line treatment.⁹⁶ However, this assumption can be regarded as conservative as any bias would be directed against escitalopram. The source of effectiveness data was clearly described for patients at all

stages of the model pathways. However, as highlighted by the authors, the use of trial data from Western Europe may not be generalisable to an Asian population. In addition, it seems counterintuitive to assume different efficacy for the same drug, depending on which comparator is used; the efficacy estimates would perhaps have been more robust had the available evidence been used to inform a MTC of all three considered treatment options. Resource-use data were clearly described and estimated from a survey of experienced clinicians in both primary and secondary care.

Benedict *et al.* 2010

Benedict *et al.*⁸⁶ conducted an economic evaluation comparing duloxetine with extended-release (ER) venlafaxine and also with mirtazapine in the treatment of MDD in Scotland from a Scottish NHS perspective. Patients with moderate to severe MDD (HAMD-17 score of ≥ 19) and patients with severe MDD (HAMD-17 score of ≥ 25) were considered separately in primary and secondary care model scenarios. SSRIs were considered as a treatment option in patients with moderate to severe MDD, whereas patients with severe MDD referred to secondary care were assumed to have previously received multiple lines of SSRI therapy and therefore the comparators were limited to ER venlafaxine and mirtazapine.

Model structure and assumptions

A cost–utility model was constructed to compare the QALYs and costs accrued with ER venlafaxine and mirtazapine with those obtained with duloxetine. The analysis was carried out over a 48-week time horizon. The time horizon was chosen to capture treatment duration as recommended by NICE¹⁶ (treatment continued for 6 months following remission and relapses within 1 year). A Markov model was used and health states were a combination of patient's disease (i.e. depressed, response, in remission, no response, relapse, recurrence) and treatment status (i.e. acute treatment, continued treatment, switch treatment and no treatment). Cycle length was 8 weeks (typical treatment duration) and patients were able to discontinue treatment in all model cycles. All patients entered the model in the acute treatment health state, from which they could experience remission, response, no response or drop out/discontinue their therapy. Patients who remitted were at risk of relapse, regardless of treatment status. Patients who responded but did not remit and remained on treatment could experience remission, response, no response or drop out in the subsequent cycle. Similarly, non-responders who did not discontinue could experience remission, response or continued non-response. Patients were eligible to switch treatments following recurrence of symptoms after initial treatment discontinuation, no response following previous response or remission, or continued non-response.

Efficacy data

The accumulation of QALYs is driven by the rate of response, remission and discontinuation throughout the acute and maintenance phases of treatment. The utility values associated with response, remission and discontinuation were derived from the European Quality of Life-5 Dimensions (EQ-5D) data of 300 European patients with MDD from the Eli Lilly HMBU trial (data on file). In moderate to severe patients, effectiveness data for acute treatment were synthesised from eight RCTs for duloxetine, two head-to-head trials for ER venlafaxine and published meta-analyses for mirtazapine and SSRIs; a class effect was assumed for SSRIs. Data for effectiveness of maintenance treatment in moderate to severe patients were taken from a head-to-head trial comparing duloxetine with ER venlafaxine. However, the absence of maintenance effectiveness data for mirtazapine and the SSRIs led to the application of a weighted average of the effectiveness of duloxetine and venlafaxine. In patients with severe MDD, the probabilities of response to acute treatment with duloxetine and ER venlafaxine were taken from two head-to-head trials. Relative differences were used to calculate the acute and maintenance probabilities for mirtazapine, and maintenance probabilities for duloxetine and ER venlafaxine.

Resource use and cost data

The perspective of the analysis was that of the Scottish NHS; therefore, direct costs are the main consideration. However, societal costs of lost productivity are accounted for in the sensitivity analysis. The number of GP visits for mental health reasons, psychiatrist visits, hospitalisations, and visits to accident and emergency departments were derived from expert panels in both primary and secondary care. The average

doses of SSRIs and mirtazapine were taken from UK market information.⁹⁷ The doses used in head-to-head trials of ER venlafaxine and duloxetine were assumed to represent standard dosing of these drugs. The authors reported that unit costs were taken from published UK sources (no reference was provided).

Summary of results

Duloxetine was associated with lower costs than venlafaxine in the severe population and lower costs than mirtazapine and venlafaxine in the moderate to severe population. More QALYs were accrued with duloxetine than any other treatment in both patient populations. However, the incremental gain was small (0.003 and 0.005 compared with venlafaxine in patients with moderate to severe MDD and severe MDD, respectively).

Consequently, duloxetine dominated (i.e. was more effective and less costly) ER venlafaxine in both patient populations and dominated mirtazapine in the moderate to severe patient population. In the moderate to severe MDD, the ICERs for duloxetine compared with mirtazapine and the SSRIs were £2353 and £6304 per QALY, respectively. The authors concluded that duloxetine displays similar efficacy and a different side effect profile to ER venlafaxine and that it represents an important treatment option for patients with MDD in the UK.

Critique

Overall, the study was considered to be high quality; the assessment of cost per QALY gained is appropriate to address the decision problem from the perspective of the Scottish NHS. All model inputs are clearly described and the analysis has several strengths, including the use of head-to-head RCT data to inform the comparison of duloxetine and ER venlafaxine and the use of (unpublished) utility values derived from RCT data. The chosen time horizon of 48 weeks seems appropriate to gather the different short- and medium-term costs and outcomes associated with the treatments of MDD. The consideration of two different patient populations is also a strength, as patients are managed differently according to their history and whether they are treated in primary or secondary care.¹⁶

For the purposes of this review, the focus is on severe patients in secondary care who are assumed to have received multiple lines of SSRI therapy. The effectiveness data used to inform this aspect of the analysis are weaker than those forming the basis of the analysis of first-line primary care patients. Although head-to-head RCT data are used to inform the acute outcomes of treatment with duloxetine and ER venlafaxine in the acute phase, maintenance therapy is based on relative differences between first and second cycle probabilities (details of calculations were not provided in the paper). Resource-use data were collected from a panel consisting of two GPs and one psychiatrist in primary care and two psychiatrists in secondary care.

No rationale for the choice of model type was provided but a Markov model may be appropriate given the cyclical nature of remission and relapse often seen in depressed patients.⁹⁸ In addition, it is not clear to what extent the progress of patients is tracked through cycles of response, relapse, treatment switch and disease recurrence. In particular, patient's progression following switch treatment is unclear. The authors were contacted to clarify this; no response has been received at the time of writing the final report.

Simpson *et al.* 2009

Simpson *et al.*⁸⁹ present a cost-effectiveness analysis of TMS in patients with moderate to severe unipolar non-psychotic MDD, defined as a MADRS score of ≥ 17 . Patients were also moderately to severely pharmacologically treatment resistant [as measured by the antidepressant treatment history form (ATHF)], having received no clinical benefit from between one and four adequate antidepressant exposures. Analysis was undertaken from a US payer and societal perspective. TMS was compared with sham TMS and pharmacotherapy as usual. Efficacy data for pharmacotherapy as usual were based on the clinical outcomes observed in the published results of the Sequenced Treatment Alternatives to Relieve Depression (STAR*D) trial.⁹⁹ The STAR*D trial⁹⁹ was a series of RCTs in outpatients with non-psychotic MDD who were eligible for medication as first-line treatment. Patients were evaluated across sequential lines of therapy (up to four lines

of therapy) and guidance was provided on how to start therapy and how to proceed if initial treatment failed.¹⁰⁰

Model description

The authors used a hybrid model, consisting of decision tree and Markov components. The decision tree component was used to simulate treatment outcomes following 6 weeks of acute treatment and a planned 3-week taper phase. Patients were classified as well (MADRS score of 0–9) or mildly (MADRS score of 10–17), moderately (MADRS score of 18–27) or severely depressed (MADRS score of >27). Following the acute treatment phase of the model, patients were disaggregated into two separate Markov components. The first accounted for the progression of patients who were well or who had mild or moderate depression following acute treatment. The second Markov component accounted for the progression of patients who were severely depressed following acute treatment; these patients were assumed to have failed at least two antidepressants and TMS therapy. Within each Markov component patients could move between the health states of well or mild, moderate or severe depression in quarterly (3-month) cycles over a 1-year time horizon.

Effectiveness data

The decision tree component of the economic model was parameterised with acute treatment outcomes and severity-specific rates of relapse were used to drive the Markov components. All TMS and sham TMS clinical efficacy parameters were derived from analysis of raw clinical effectiveness data, provided by the manufacturer of TMS (Neuronetics, Inc., Malvern, PA, USA). Patients with severe depression following acute TMS treatment were assumed to have the same potential to benefit from further antidepressant therapy as patients in levels 3 and 4 of the STAR*D study⁹⁹ (i.e. patients who had failed on two or three previous therapies). Patients who were well or had mild-to-moderate depression following acute TMS treatment were assumed to have the same potential to benefit from further antidepressant therapy as patients in levels 2 and 3 of the STAR*D study⁹⁹ (i.e. patients who had failed on one or two previous treatments). In addition, a subgroup of TMS patients who had failed only one previous therapy in their current episode of depression were compared with level 2 patients of the STAR*D trial.⁹⁹ QoL weights were obtained from a study by Revicki *et al.*⁹¹

Resource use and cost data

Health-care resource utilisation was estimated for each health state using data from the results of self-report questionnaires used in the Neuronetics trials. The questionnaire covered lost productivity, health-care utilisation and costs, and caregiver support. Unit cost data were taken from the 2004 Medicaid billing database for patients with depression or the Neuronetics studies; all costs were inflated to 2006 values. The overall cost associated with each treatment regimen was dependent on the composition of the patient population, treatment efficacy and costing assumptions used. Analyses were carried out from both the payer and societal perspective.

Results

Mean annual costs (excluding lost productivity) of STAR*D patients varied from US\$4379 to US\$26,546 for patients who responded to initial therapy and non-responders, respectively; non-responders often required hospitalisation and multidrug treatment. Similarly, mean annual costs for TMS patients (excluding the cost of TMS) ranged from US\$3683 for responders to US\$26,599 for non-responders. The cost of TMS was estimated to be US\$300 per treatment.

Incremental cost-effectiveness results were presented for the following comparisons:

- TMS vs. sham TMS (based on clinical data from a RCT)
- TMS vs. sham in patients with ATHF = 1 (based on subgroup analysis of a RCT)
- TMS vs. pharmacotherapy as usual (based on an open-label TMS study and the level 2 and 3 outcomes of STAR*D⁹⁹)
- TMS vs. pharmacotherapy as usual in patients with ATHF = 1 (based on a subgroup of an open-label TMS study and the level 2 outcomes of STAR*D⁹⁹).

No results were presented for the more severe patient population. TMS resulted in cost savings (i.e. dominated) of US\$746 and US\$2243 compared with pharmacotherapy as usual in the level 2/3 STAR*D patients and the level 2 STAR*D patients, respectively. These cost savings increased to US\$7243 and US\$9844 when lost productivity and increased caregiver costs, respectively, were included in the analyses. The comparison of TMS with sham TMS (i.e. no treatment) resulted in ICERs of US\$36,551 and US\$29,556 in all randomised patients and patients with ATHF scores of 1, respectively. The inclusion of lost productivity and caregiver costs into the analyses resulted in an ICER of US\$3544 for TMS compared with sham in all eligible patients and cost savings of US\$5092 in patients who have failed only one previous antidepressant therapy.

Critique

The study presented by Simpson *et al.*⁸⁹ had many strengths, not least of which was the large volume of acute treatment data derived from RCTs used to inform the model transitions. Furthermore, the length of follow-up of RCT patients was sufficient to address questions over the durability of treatment effect. However, the description of which method was used to extrapolate 6-month follow-up data to 1 year was insufficient to determine its methodological robustness. The comparison of TMS with the STAR*D study⁹⁹ was largely reflective of clinical practice. However, the expected difference in costs and outcomes between TMS and distinct treatment regimens or sequences was not quantified. The use of individual patient surveys to collect health-care resource utilisation data is another particular strength of this analysis. The use of QoL data from Revicki *et al.*⁹¹ is consistent with many studies in this disease area. However, adverse events and treatment-related QoL were not considered. The absence of adverse events is the main weakness of this study, along with insufficient reporting of gains in QALYs and the inconsistent reporting of ICERs between the summary table and the text.

Quality-of-life literature review

Introduction

This literature review was carried out to identify utility values associated with depression health states in a patient population with TRD. As in the review of the economic literature, it was expected a priori that, if available, QoL literature on TRD would be limited. In addition, it was considered that the health-related quality of life (HRQoL) associated with TRD was unlikely to be different to that associated with depression in general. Therefore, a decision was made to expand the remit of the search to consider studies of depression in general rather than limiting to studies in TRD. Hence, the primary objectives of this review were to identify health-state utility values (HSUVs) associated with different levels of depression. The secondary objective of this review was to identify any issues that positively or negatively impact upon the QoL in depressed patients.

The review of the QoL literature identified five HSUV studies that addressed the primary objective of the review: to retrieve utility values associated with different levels of depression. A further 12 studies^{30,43,46–53,59} were identified that addressed the secondary objective: to understand the scope of issues that affect quality of life in the TRD patient population. The sections that follow provide an overview and quality assessment of the HSUV studies identified as addressing the primary objective of this review; the rationale for the utility values used in the economic model; and an overview of the QoL studies identified as addressing the secondary objective of this review. The quality assessment of HSUV studies is based on quality assessment criteria outlined by NICE's decision support unit: Technical Support Document 9 2011 (TSD 9 2011).¹⁰¹ As recommended by the decision support unit, factors considered when evaluating quality included selection and recruitment of respondents, inclusion and exclusion criteria and whether or not the study included a description of the baseline characteristics of the population from which values were derived. Response rates of the measures used to derive the HSUVs and loss to follow-up were also evaluated.

Systematic literature search and selection process

As stated above, a systematic literature search was carried out to identify utility values associated with depression health states and to identify any issues that positively or negatively impact on the QoL of depressed patients.

Literature search terms and strategies used to identify cost-effectiveness studies relevant to the decision problem were based on validated search strategies developed by Haynes *et al.*⁸⁴ and Dickersin *et al.*⁸⁵ Multiple databases encompassing medical and economic literature were searched to maximise the potential of capturing relevant studies. Searches were carried out in the following databases:

- EMBASE (for the period 1988 to August 2011)
- MEDLINE, including MEDLINE In-Process & Other Non-Indexed Citations (for the period 1950 to August 2011)
- CENTRAL (from inception to August 2011)
- HTA database (from inception to August 2011)
- NHS EED accessed via Wiley Online Library (for the period 1999 to August 2011)
- PsycINFO (from inception to August 2011).

Details of the search strategy are provided in *Appendix 7*. The search terms used covered condition, population and intervention; no country or language restrictions were applied. All references were exported to the Reference Manager bibliographic database and deduplicated. In addition, the reference lists of identified systematic reviews were hand-searched for additional references. Inclusion and exclusion criteria specific to the QoL literature review were developed (*Table 8*).

The initial search identified 352 papers, of which 114 were duplicates. One reviewer (LN) carried out the first appraisal of the title and abstract of the 238 studies (level one screening) for potential inclusion. At this stage, 205 papers were excluded.

Further assessment of included abstracts (level two screening) was carried out by a second reviewer (NT) and 13 abstracts were excluded. The full-text publications of 29 studies identified as potentially relevant by both reviewers were ordered (including nine studies identified through bibliographic hand-searching). The papers were independently assessed for final inclusion by two reviewers (LN and NT) using the criteria presented in *Table 8*: 12 studies were excluded at this stage. *Figure 12* summarises the selection process.

Overview of included health-state utility valuation studies

The five HSUV studies identified in the QoL literature review reported utility values for depression severity and/or treatment response. Two studies^{102,103} reported utility by depression severity, one study¹⁰⁴ reported utility associated with treatment response and two studies^{91,105} considered both depression severity and treatment response. Most studies were carried out in North America (three in Canada^{91,102,103} and one in the USA;¹⁰⁴ the remaining study was carried out in France¹⁰⁵). The studies were published between 1998 and

TABLE 8 Inclusion and exclusion criteria applied in the review of the QoL literature

Criteria	Inclusion	Exclusion
Type of study	Any study type	None
Population	People with TRD, MDD or any severity of depression	People with bipolar disorder or psychotic conditions
Geographical location	Publications from any country	None
Interventions	Any	None
Outcomes of interest	Utility scores, method of elicitation/valuation General QoL issues	None

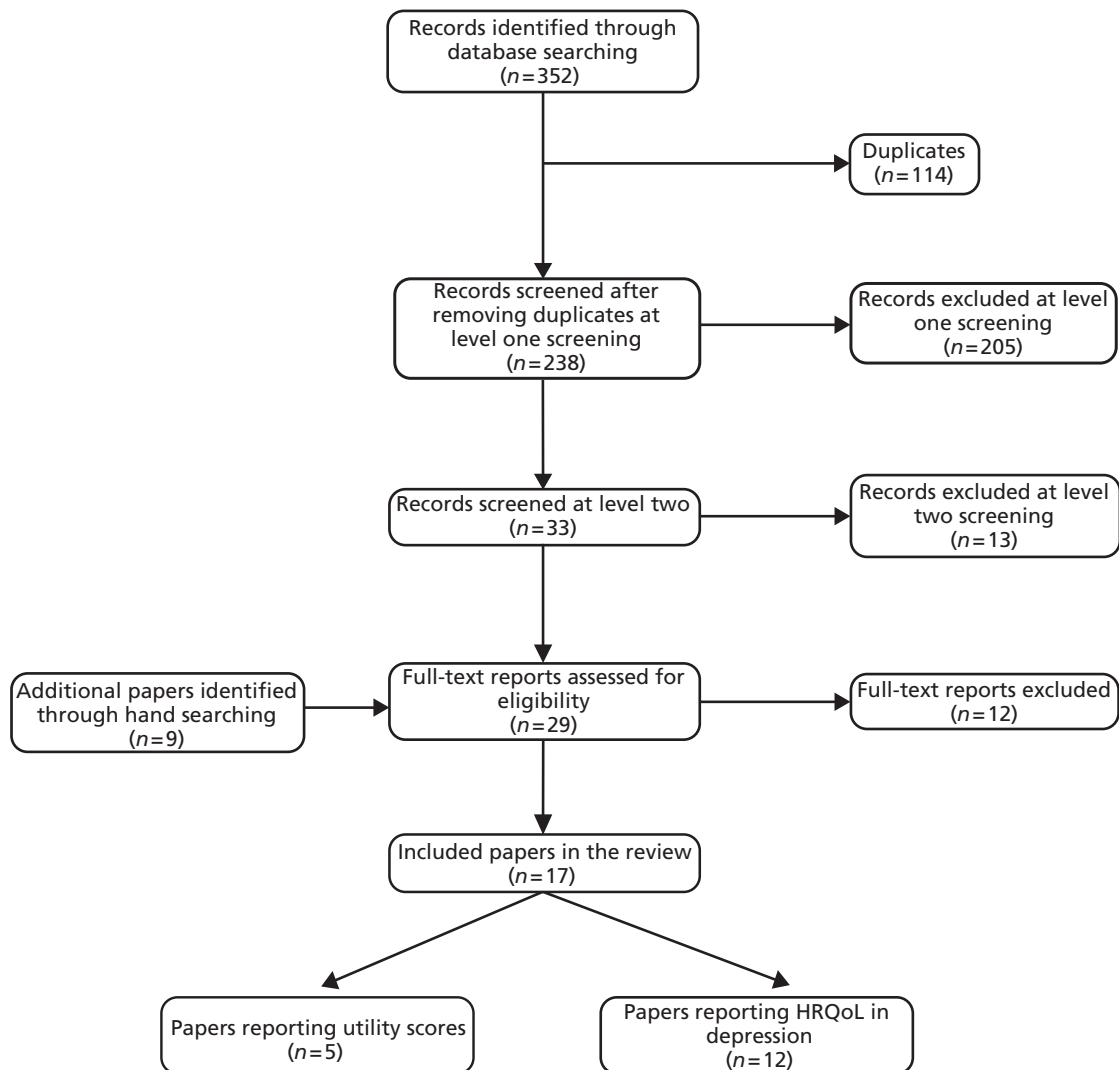


FIGURE 12 Flow diagram of studies in systematic review to identify HSUVs.

2004 and sample sizes ranged from 58 to 250. The studies each used a different tool to evaluate health-state utility: the tools included the EQ-5D questionnaire, the self-administered Quality of Well-Being Scale (QWB-SA), the SF-36, and the McSad instrument. The mean age of the study participants was < 50 years and more than two-thirds of study participants were women. *Table 9* below summarises the five included studies.^{91,102–105}

Narrative review of included studies (critique of identified quality-of-life studies)

Revicki *et al.* 1998

Revicki *et al.*⁹¹ used standard gamble (SG) techniques to elicit utility values from 70 Canadian (57% of participants) and US (43% of participants) primary care patients. Patients had MDD (DSM-III-revised) with mean HAMD score of 11.65 (SD 8.2). Patients were either currently receiving treatment or had completed an antidepressant treatment in the past 2 months. The mean age of participants was 42 years (SD 11 years) and 77% were female. Health status was measured using the SF-36. Patients were presented with 11 hypothetical depression-related health states based on vignettes for depression health states and treatment. The vignettes were informed by literature and three psychiatrists experienced in treating depression. Utility was generated by applying a structured SG interview for each hypothetical health state.

TABLE 9 Summary of utility values reported in all included HSUV studies

Author, year, country	Sample size	Patient population ^a	Instrument (valuation)	Utility results
Pyne <i>et al.</i> (2003), ¹⁰⁴ USA	58	Patients with MDD (DSM-IV criteria)	QWB-SA	<p><i>Responders</i></p> <p>Baseline: 0.43</p> <p>4 weeks: 0.54</p> <p>4 months: 0.63</p> <p><i>Non-responders</i></p> <p>Baseline: 0.41</p> <p>4 weeks: 0.46</p> <p>4 months: 0.43</p>
Revicki <i>et al.</i> (1998), ⁹¹ Canada	70	Patients with unipolar depression seen in US and Canadian primary care	The health states were hypothetical, based on vignettes	<p>Mean (SD)</p> <p><i>Current health state (following at least 8 weeks of antidepressant treatment)</i></p> <p>0.74 (0.22)</p> <p><i>Severe depression untreated</i></p> <p>0.30 (0.28)</p> <p><i>Moderate depression</i></p> <p>Nefazodone: 0.63 (0.23)</p> <p>Fluoxetine: 0.63 (0.19)</p> <p>Imipramine: 0.55 (0.03)</p> <p><i>Mild depression</i></p> <p>Nefazodone: 0.73 (0.21)</p> <p>Fluoxetine: 0.70 (0.20)</p> <p>Imipramine: 0.64 (0.20)</p> <p><i>Remission: maintenance treatment</i></p> <p>Nefazodone: 0.83 (0.13)</p> <p>Fluoxetine: 0.80 (0.15)</p> <p>Imipramine: 0.72 (0.17)</p> <p><i>Remission: no treatment</i></p> <p>0.86 (0.16)</p>
Schaffer <i>et al.</i> (2002), ¹⁰³ Canada	75	A mixture of people with current depression, previous depression (according to MDD DSM-IV criteria) and healthy people	The health states were hypothetical, based on vignettes	<p>Mean (SD)</p> <p><i>Currently depressed patients</i></p> <p>Severe: 0.31 (0.31)</p> <p>Moderate: 0.51 (0.34)</p> <p>Mild: 0.59 (0.33)</p> <p><i>Previously depressed patients</i></p> <p>Severe: 0.47 (0.34)</p> <p>Moderate: 0.67 (0.36)</p> <p>Mild: 0.79 (0.28)</p>

TABLE 9 Summary of utility values reported in all included HSUV studies (continued)

Author, year, country	Sample size	Patient population ^a	Instrument (valuation)	Utility results
				<i>Healthy individuals</i>
				Severe: 0.46 (0.28)
				Moderate: 0.69 (0.29)
				Mild: 0.80 (0.21)
				<i>All participants</i>
				Severe: 0.43 (0.31)
				Moderate: 0.64 (0.33)
				Mild: 0.75 (0.28)
Bennett <i>et al.</i> (2000), ¹⁰² Canada	105	Patients currently in remission, with history of unipolar depression in the past 2 years	McSad (SG and VAS)	Mean (95% CI) <i>Self-reported health-state</i> 0.79 (0.74 to 0.83)
				<i>Temporary states (6 months' duration)</i>
				Mild: 0.59 (0.55 to 0.62)
				Moderate: 0.32 (0.29 to 0.34)
				Severe: 0.09 (0.05 to 0.13)
				<i>Chronic states (lifetime duration)</i>
				Severe: 0.04 (0.009 to 0.07)
Sapin <i>et al.</i> (2004), ¹⁰⁵ France	250	Patients with new episodes of MDD (DSM-IV) seen in primary care	EQ-5D (TTO)	Mean (SD) <i>Disease severity health states (after 8 weeks of treatment)</i> Mild: 0.74 (0.19) Moderate: 0.44 (0.27) Severe: 0.30 (0.27)
				<i>Treatment response health states (after 8 weeks of treatment)</i>
				Remission: 0.85 (0.13)
				Response: 0.72 (0.20)
				Non-response: 0.58 (0.28)

NDC, non-directive counselling; TTO, time trade-off; VAS, visual analogue scale.

^a Patients enrolled in the study. Details of who provided the health-state descriptions and who provided the valuations are provided in the narrative section below.

Health states were a combination of depression severity (mild, moderate or severe) and antidepressant treatment [nefazodone (Serzone[®], Bristol-Myers Squibb), fluoxetine or imipramine]. Health-state profiles were framed as enduring for 1 month and each profile contained descriptions of symptom severity, functioning and well-being, and side effects from treatment. The authors concluded that there were significant differences between the mean utility scores generated for the hypothetical health states, depending on a patient's current severity of depression; more severely depressed patients (HAMD score of ≥ 15) and patients experiencing three or more side effects gave lower scores for hypothetical remission states than patients currently in remission. Although the study was of good quality, despite reporting the number of patients with missing or incomplete utility data, the authors did not state how they had dealt with missing or incomplete data.

Bennett *et al.* 2000

Bennett *et al.*¹⁰² used a disease-specific measure, the McSad instrument, to estimate utility scores for a cross-sectional sample of 105 Canadian patients. Patients were currently in remission and had experienced at least one episode of major unipolar depression in the previous 2 years. The mean age of respondents was 41.7 years (SD 8.7 years) and 74% were female. The authors reported that McSad is a direct utility measure of depression health states that uses a combination of rating scale and SG techniques to obtain utility values. Based on the DSM-III-revised criteria for major unipolar depression, McSad assesses six dimensions: emotion; self-appraisal; cognition; physiology; behaviour; and role function. Patients provided utility values for four depression health states, including three hypothetical 'clinical marker' health states of untreated depression (mild, moderate and severe) and current self-reported health. Utility scores for the three clinical marker health states were framed as enduring for 6 months, and the chronic states (self-reported and severe depression) were measured assuming lifetime duration. The authors concluded that depression is 'associated with poor HRQoL'; moderate depression generated utility values that were lower than those reported for patients who are blind, deaf and dumb. However, the authors also stated that these findings may not be generalisable, as the extent to which the sample is representative is unknown. The study appears to be of good quality, as it satisfies most of the NICE TSD quality criteria.¹⁰¹

Schaffer *et al.* 2002

Schaffer *et al.*¹⁰³ elicited utility scores from a mixed population in Canada, consisting of 40 patients with MDD (meeting DSM-IV criteria) and 35 healthy people. The mean age of participants was 42.4 years (SD 11.4 years) and 67.5% were female. Utility values were assigned using SG techniques. Participants were presented with 10 profiles describing different symptoms of depression and three profiles describing depression of a mild, moderate or severe nature. The individual symptom profiles contained five descriptive statements, which were derived from depression rating scales and interviews such as the Beck Depression Inventory (BDI), DSM-IV, HAMD and MADRS. Somatic symptoms (e.g. low energy and decreased appetite or sleep) were rated more highly than psychological symptoms (e.g. suicidal ideation, depressed mood and anhedonia), with suicidal ideation having the lowest utility score. People with current depression generally assigned lower utility scores to profiles than did healthy people or participants with a history of depression. In addition, no significant difference in utility was found between people with a history of depression and healthy individuals. The authors suggested that the presence of current depression may affect utility scores and that this could be an important consideration in patients with primary diagnoses other than depression. However, the authors also highlighted the limitations to this study of small sample size and the use of SG techniques in depressed patients, who as a consequence of their illness may not be as risk averse as people who are not depressed. Overall, the study was of good quality, with respondent selection, recruitment, and inclusion and exclusion criteria clearly described. There was no discussion around patients lost to follow-up but the results suggest that all patients enrolled were followed up for the duration of the study.

Pyne *et al.* 2003

Pyne *et al.*¹⁰⁴ used the QWB-SA in a prospective observational study. Fifty-eight US patients with a current diagnosis of MDD were monitored over 16 weeks to assess the relationship between QWB-SA scores and depression severity. Consistency was assessed by comparing scores obtained with the QWB-SA and those obtained from the interviewer-administered version (interviewer-QWB). The mean age of participants was 45.7 years (SD 10.3 years), 83% of participants had unipolar depression and 17% had bipolar depression, and 78% were male. The QWB-SA is a generic HRQoL instrument with five domains (symptoms, self-care, mobility, physical function and performance of usual activity). The output of QWB-SA is a quality-adjusted index score of between 0 (death) and 1 (perfect health) derived from scores in the five domains.¹⁰⁶ QWB-SA scores improved during follow-up for treatment responders (defined by a 50% reduction in HAMD-17 scores from baseline) but did not improve for non-responders. The authors concluded that QoL was associated with depression severity, with responders having a better QoL than non-responders. The study appears to be of good quality with clearly described methods.

Sapin *et al.* 2004

Sapin *et al.*¹⁰⁵ examined the impact of MDD on patients' HRQoL using data collected from a multicentre non-comparative prospective cohort study in patients with a new episode of MDD. Patients ($n = 250$) were recruited in the French primary care setting and followed for 2 months. The mean age of participants was 44.2 years (SD 14.1 years) and 72% were female. Patient preferences were elicited through the use of the EQ-5D questionnaire. The resultant health states were assigned utility values derived from a large UK survey using time trade-off (TTO) techniques.¹⁰⁷ EQ-5D utility scores were reported by depression severity [defined by the Clinical Global Impression of Severity scale (CGI-S)], and by clinical response (defined by MADRS scores) at follow-up. Patients were classed as remitters, responders or non-responders. Remitters were patients with MADRS scores of ≤ 12 , and responders were patients with a reduction in their MADRS score of $\geq 50\%$ from baseline score; all other patients were classified as 'non-responders'. The difference in utility scores among remitters, responders and non-responders was statistically significant ($p < 0.001$) from baseline at 4 and at 8 weeks. In addition, the authors also found that sex and age did not influence utility value: there was no statistically significant difference in utility scores between men and women and no difference by age. Overall the study was well described and of good quality, and satisfies the quality assessment criteria outlined in the TSD by NICE's decision support unit.¹⁰¹

Utility data used in the model

There is ongoing debate around the optimum approach for measuring patient outcomes for use in economic evaluation and decision making.¹⁰¹ Areas of uncertainty include which health status instrument to use [EQ-5D, Health Utilities Index (HUI) or Short Form questionnaire-6 Dimensions (SF-6D)]; which valuation technique [e.g. TTO, SG or visual analogue scale (VAS)] to apply; and whose preferences (patients, clinicians or the general public) to consider. However, NICE's methods guide¹⁰⁸ recommends that:

- (a) Health status should be reported by the patients experiencing the condition.
- (b) The values placed on changes in health should come from the UK general population using a choice-based method.
- (c) EQ-5D is the preferred measure of HRQoL in adults.

In line with these recommendations, one of the HSUV studies identified in the review of the QoL literature, Sapin *et al.* (2004),¹⁰⁵ met the above criteria and was considered as the source of utility for the model. Utility was reported by level of treatment response (remission, response and non-response) and by severity of depression (mild, moderate and severe). The study was carried out in French patients but used UK general public valuation scores. The model assumes that at baseline everyone had the utility of severe depression, as all patients in this model have TRD.

Overview of included quality-of-life studies

The inclusion criteria for this review were broad, and studies examining HRQoL in depression of any severity (not limited to TRD) were considered in addition to HSUV studies. This was to identify any issues that positively or negatively impact on the QoL of depressed patients (these issues were considered likely to be similar in depressed patients, regardless of level of treatment resistance). Twelve studies considering HRQoL were identified that covered a wide range of issues relevant to depression, including the impact on HRQoL of treatment (drug dosage and duration of treatment), physical illness, exercise and depression severity. Ten of the 12 studies were conducted in the USA. Of these 10 studies, four measured health status using the SF-36,¹⁰⁹⁻¹¹² three used the Quality of Life Enjoyment and Satisfaction Questionnaire (Q-LES-Q),^{67,69,113} one used both SF-36 and Q-LES-Q,¹¹⁴ and one used the World Health Organization Quality of Life Assessment (WHOQOL-BREF).¹¹⁵ Finally, one study used both the SF-36 and the Late-Life Function and Disability Instrument (LLFDI).¹¹⁶ A further study using the WHOQOL-BREF was carried out in Brazil¹¹⁷ and a study using the Sickness Impact Profile (SIP) was carried out in the Netherlands.¹¹⁸

The patient populations examined varied across studies; however, two studies^{69,110} included patients with TRD as defined in this review (failed on at least two adequate antidepressant treatments in the current

episode of depression). Two studies^{109,117} assessed outcomes in patients with unipolar depression. The remaining studies^{67,111–118} included patients meeting DSM-IV criteria for MDD. A range of interventions were considered in the identified studies: eight studies^{69,110,111,113,115–118} considered antidepressant treatment (SSRI/SNRI/tricyclic antidepressant); two studies^{67,112} considered augmentation of antidepressant therapy with risperidone; one study¹¹⁴ considered the augmentation of antidepressant therapy with physical exercise; and the remaining study¹⁰⁹ considered any therapeutic regimen that had been agreed by patients and their psychiatrist. All studies reported an improvement in QoL with treatment.

There is some evidence that the impact of treatment on QoL is sustained and a longer duration of treatment results in further improvement. A 2-year maintenance study by Trivedi *et al.*¹¹³ assessed the psychosocial outcomes in patients with recurrent MDD. The patients were responders to ER venlafaxine during acute and continuation phase and at 1- and 2-year maintenance periods. The authors found that the longer a patient continues treatment, the higher their QoL. The study found that people who responded to treatment and switched to placebo had worse QoL than those who remained on treatment. The authors also found that patients with greater severity of depression have worse QoL across multiple domains and concluded that improvement in depressive symptoms would be reflected in assessment of functioning.

Another study by Karp *et al.*¹¹⁶ assessed the correlation between disability depression severity and QoL in patients with or without depression. The study used the LLFDI, which is defined by the authors as a measure of instrumental activity of daily living, personal role and social role function. The authors found that there was a correlation between disability and depression severity measured on the HAMD-17 scale and that antidepressant treatment improved functional ability. The study also demonstrated that those patients who stayed longer on treatment continued to improve.

In addition, a prospective multicentre observational study by Dunner *et al.*¹¹⁰ in people with TRD who were severely depressed and receiving treatment as usual [any therapeutic regimen agreed to by the treating physician and the patient (i.e. drugs or ECT)] found that, after 12 months of treatment, 48% of the patients reported that their QoL had not changed. After 2 years of treatment, the proportion of patients reporting no change in QoL fell slightly to 42%. The proportion of patients reporting an improvement in QoL was 30% and 36% at 12 months and 24 months' follow-up, respectively. Therefore, based on the results from the studies by Trivedi *et al.*¹¹³ and Dunner *et al.*¹¹⁰ it could be inferred that patients should be kept on treatment as long as possible to maximise the HRQoL gains. There is, however, some evidence from relapse prevention studies in depression that indicates a higher rate of relapse in patients who cease therapy following response.¹⁵ It may be that patients in Trivedi *et al.*¹¹³ and Dunner *et al.*¹¹⁰ who ceased therapy or switched to placebo experienced a 'rebound' effect that resulted in a lower QoL than in patients who remained on therapy.

About half of the studies reported that the physical domain of QoL for different instruments (SF-36, SIP, Q-LES-Q and WHOQOL-BREF) did not improve with treatment (antidepressants or antidepressant plus risperidone).^{109,110,112,113,118} However, one study by Carta *et al.*¹¹⁵ assessing physical activity as an adjunctive therapy in women with MDD found that the physical component of the WHOQOL-BREF improved significantly in women who did physical activity compared with women who did not. There were no differences in other domains (relationships, environment and psychological) between those who did physical activities and those who did not. These findings suggest that physical activity may be used as adjunctive therapy to antidepressants to improve the overall QoL of people suffering from MDD.

A study by Small *et al.*¹⁰⁹ in elderly patients (age > 60 years) with unipolar depression assessed the impact of physical illness on QoL. The indicators for physical illness were number of current chronic and historical illnesses and these indicators were found to influence QoL measures. QoL was measured using the SF-36, and depression levels were the same between patients with physical illness and those who did not have physical illness. The commonly reported chronic illnesses were joint diseases, cardiovascular disease, allergies and gastrointestinal disease. Frequently reported historical illnesses were surgical procedures, gastrointestinal disorder, accidental injuries and cardiovascular disease. The authors concluded that current and previous

physical illness is associated with poor QoL, especially physical and mental functioning in older patients with depression. Therefore, based on the findings of the studies by Carta *et al.*¹¹⁵ and Small *et al.*,¹⁰⁹ physical and psychological health may be considered as relatively independent factors of QoL.

In conclusion, there is evidence to suggest that treatments for depression are effective in improving the QoL of patients, particularly if treatment is sustained. In addition, the QoL evidence base indicates that pharmacological treatment of depression does not result in any improvement of physical outcomes, and physical therapies do not result in any improvement of psychological symptoms. This, in turn, suggests that the physical and psychological domains of QoL are independent of each other in determining the overall QoL of a depressed patient.

Chapter 5 De novo economic analysis

Introduction

As many as two-thirds of patients with major depression will either not respond to, or have a suboptimal response to, first-line treatment with antidepressants.¹⁶ After an inadequate response to at least one antidepressant treatment, current NICE depression guidance¹⁶ recommends that:

If a person with depression is informed about, and prepared to tolerate, the increased side effect burden, consider combining or augmenting an antidepressant with lithium or an AAP, such as aripiprazole, olanzapine, quetiapine or risperidone or another antidepressant, such as mirtazapine or mianserin.

The focus of this report is the comparative clinical effectiveness and cost-effectiveness of augmentation of SSRI therapy with lithium or with an AAP in a TRD population. For the purposes of this report, TRD is defined as those patients with inadequate response to two or more adequate trials of antidepressant therapies, as specified in the NIHR HTA final protocol.²⁴

To facilitate the effective and efficient allocation of health-care resources, it is necessary to quantify and compare the economic benefits of the two key augmentation strategies of lithium and AAPs. A review of the cost-effectiveness literature carried out as part of this report (see *Chapter 4, Economic literature review*) did not identify any analyses comparing the two augmentation strategies in a TRD population. Therefore, a de novo economic analysis was developed to estimate the costs, consequences and relative cost-effectiveness of each augmentation strategy over a 1-year time horizon. The model was constructed in Microsoft Excel™ (Microsoft Corporation, Redmond, WA, USA) and used probabilistic analysis of a cohort of 2000 TRD patients to estimate the expected cost-effectiveness of each strategy.

Model overview

A hybrid economic model was constructed to simulate the clinical and economic consequences of augmenting an SSRI with either lithium or an AAP in the treatment of TRD. The model considered outcomes from the perspective of the NHS. This modelling approach was inspired by the model developed by Simpson *et al.*⁸⁹ discussed in the review of the cost-effectiveness literature in *Chapter 4* (see *Narrative review of included studies*).

The hybrid model facilitates capturing the granularity of the acute treatment phase while simultaneously accounting for patient progression within 1 year (i.e. discontinuation, relapse, remission and treatment response). It is likely that each treatment considered will have a differential impact on costs and effects over a longer period of time than 1 year. However, as a result of the paucity of long-term follow-up data (the effectiveness data identified in the clinical review had follow-up of, at most, 8 weeks) and, in line with published literature,⁸⁶ the model adopted a pragmatic time horizon of 1 year. Extrapolation beyond 1 year would increase model uncertainty. The economic model consists of two distinct components:

1. decision tree
2. Markov model.

A decision tree was used to simulate treatment outcomes during the acute treatment phase, which, in accordance with available clinical data, is defined as 8 weeks of treatment. All patients will be initiated on lithium or an AAP (as augmentation of SSRI therapy). The outcomes of the acute treatment phase (with or

without discontinuation) are response, non-response and remission. The schematic representation of the acute treatment phase of the model is shown in *Figure 13*.

Following on from the acute treatment phase, all patients are transitioned into the Markov component of the model. The Markov component of the model captures the maintenance phase, which includes 6 months of maintenance therapy and 4 months of follow-up (where patients may remain on or cease therapy depending on the level of response achieved). The health states captured in the maintenance phase of the model were response (with and without discontinuation), remission (with and without discontinuation), non-response (all patients in non-response are assumed to have discontinued therapy) and relapse (*Figure 14*). As a result of the lack of long-term comparative clinical data, transitions between health states within the Markov component of the model are assumed to be independent of treatment. Patients move between health states for 10 months, in cycles of 2 months, chosen to represent the usual length of a treatment course. The cycle length was considered to be small enough not to require a half-cycle correction. The absence of a half-cycle correction may result in the underestimation of the amount of relapse, discontinuation and response experienced by patients. However, given that the Markov phase of the model is treatment independent this is unlikely to bias the results of the model.

To ensure the model reflects the management of unipolar depression in the UK, two clinical advisors (both of whom are practising psychiatrists with experience of managing TRD in either primary or secondary care)

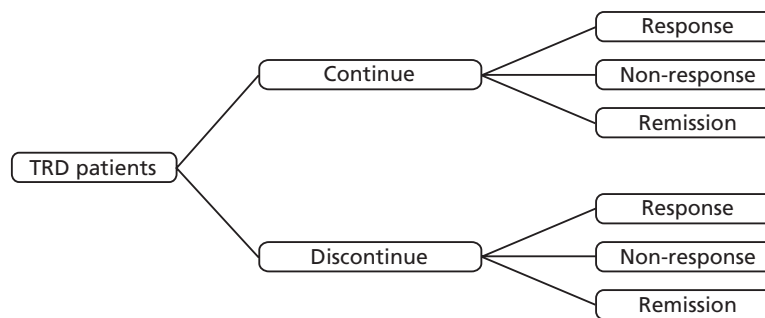


FIGURE 13 The decision tree component of the model for the cost-effectiveness of augmentation of SSRIs with lithium or an AAP during the acute treatment of TRD.

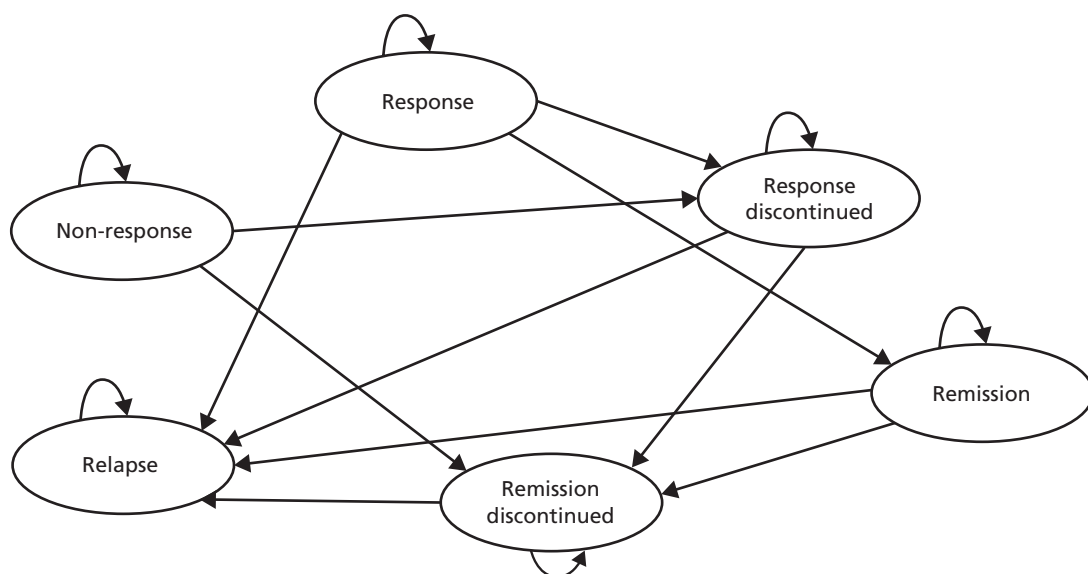


FIGURE 14 The Markov component of the model for the cost-effectiveness of augmentation of SSRIs with lithium or an AAP during maintenance-phase treatment in TRD.

were consulted to validate model assumptions throughout the model development process. Information from experts was obtained through teleconferences and standardised questionnaires.

The model considers direct treatment costs incurred by the NHS. These include costs of the pharmacological treatments, HCPs, hospitalisation and monitoring. Societal costs were not included within the model. All costs used were 2011 costs: no discounting was applied because the time horizon was no longer than 1 year. Effectiveness was measured in terms of QALYs as per the NICE reference case.¹⁰⁸ The model did not consider dose titration/escalation as, following advice from clinical experts, it was assumed that patients were on maximum tolerated doses. In addition, clinical experts advised that adverse events are unlikely to be treated; patients are more likely to discontinue their therapy. Therefore, the model does not explicitly consider the costs of treating adverse effects or the disutilities associated with adverse effects.

Model structure

Patients entered the model following initiation of augmentation therapy (lithium or an AAP) and progressed through the decision tree component of the model, representing the acute phase of treatment. Following initiation, patients were exposed to the risk of discontinuation. Discontinuation can occur for any reason; however, lack of efficacy and treatment-related adverse effects are frequently reported as the main causes of discontinuation.¹¹⁹ Expert opinion suggests that discontinuation due to side effects is usually instantaneous, whereas discontinuation owing to lack of efficacy may take place between 2 and 6 weeks after initiation of augmentation treatment. For simplicity, and owing to the way discontinuation was reported in the studies, the model does not distinguish between discontinuation due to lack of efficacy and discontinuation due to treatment-related adverse effects. Those who discontinue their treatment are expected to do so at week 4 in accordance with discontinuation owing to lack of efficacy; this is a simplifying assumption agreed by clinical experts.

Following 8 weeks of treatment, patients entered one of the following health states: 'remission', 'response' or 'non-response'. The review of the evidence available on clinical effectiveness (see *Chapter 3, Quality assessment*) indicated that symptom severity was predominantly assessed using the MADRS. Therefore, in the economic model, the health states of 'response' and 'remission' were defined using this scale. Response, which can be thought of as a period during which 'an improvement of sufficient magnitude is observed such that the individual is no longer fully symptomatic',¹²⁰ was defined as the proportion of patients who, at 8 weeks, had a reduction in their MADRS score of at least 50% (from baseline). Remission is a 'period during which an improvement of sufficient magnitude is observed that the individual is asymptomatic'.¹²⁰ Remission was defined differently in the identified trials, with some trials defining remission as a total MADRS score of ≤ 8 for two consecutive visits and others defining it as a total MADRS score of ≤ 10 at the end of the study. Based on discussions with clinical experts, remission in the model is defined as a MADRS total score of ≤ 10 at the end of the acute treatment phase.

Patients who completed acute therapy but did not enter the health states of 'remission' or 'response' moved into the health state of 'non-response', defined as those patients not achieving a 50% reduction in MADRS score at the end of the acute treatment phase. Of the patients who do not complete acute therapy owing to discontinuation, a proportion is assumed to clinically improve. Therefore, amongst people who discontinue, some will respond and some will enter remission (clinical expert opinion). Similar to patients who complete acute therapy, patients who discontinue and do not enter the health states of 'remission' or 'response' enter the health state of 'non-response'.

On completion of the acute treatment phase, all patients, irrespective of whether they have responded or discontinued their acute therapy, are moved into the maintenance phase of the model. The maintenance phase is assessed using the Markov component of the model and includes 6 months of maintenance therapy and 4 months of follow-up (on which patients may remain or cease treatment, depending on the level of response achieved). Therefore, depending on their health state at the end of the acute treatment phase,

patients enter the Markov component of the model in one of five health states: 'remission', 'remission discontinue', 'response', 'response discontinue' or 'non-response'. In addition to these health states, the Markov model includes the health state of 'relapse'. A patient in the fully symptomatic health state (MADRS score of > 15) after having achieved remission or response is considered to have relapsed. Thus, a relapse represents the return of the symptoms of a still ongoing but symptomatically suppressed episode.¹²⁰

Within the Markov model, patients may remain 'in state' or move to other health states, depending on the risks associated with their current health state. Patients who are in remission are exposed to the risk of relapse or discontinuation (of their maintenance therapy). Patients in the response health state may transition to the health states of 'remission' or 'response discontinuation'. Patients in the health state of 'remission discontinuation' may relapse, whereas those in 'response discontinuation' may transition into 'relapse' or 'remission discontinuation'.

Patients who have not responded to their acute treatment (irrespective of whether or not an acute therapy course was completed) are assumed to receive no further benefit from augmentation therapy and are assigned to a standard package of care (full details of the standard package of care are given below; see *Costs associated with each health state*). However, a proportion of non-responders are assumed to clinically improve, i.e. some will respond and some will enter remission (clinical expert opinion). Therefore, patients in the health state of 'non-response' may move to the health states of 'remission discontinuation' or 'response discontinuation'.

Similar to the state of 'non-response', people who relapse are assumed to receive no further benefit from augmentation therapy and are assigned a standard package of care. This simplifying assumption was used in place of explicitly modelling follow-on patient care pathways, as these are diverse and there is a paucity of data to inform them. However, unlike the health state of 'non-response', the health state of 'relapse' is an absorbing state, as patients entering this health state do not experience any further transitions and remain 'in state' for the duration of the model. In contrast to non-response, relapse is assumed to be the return of symptoms that had previously been suppressed. Therefore, the probability of entering response or remission following relapse is likely to be considerably different from the probability of entering response or remission from a state of non-response. No data were identified in the literature to inform these transitions. However, the Markov component of the model uses transition probabilities that are independent of treatment and it is, therefore, unlikely that assuming relapse is an absorbing state would bias the model over the short time horizon considered.

Effectiveness data

Acute treatment phase

Patient outcomes following the decision tree component of the model are dictated by the clinical efficacy of and the proportion of patients completing their acute augmentation therapy. In this section, discussion focuses on the:

- acute efficacy associated with each augmentation therapy
- rate of discontinuation associated with each therapy
- likelihood of clinical improvement of symptoms after discontinuation from acute therapy.

Acute efficacy

The acute efficacy of augmentation therapy with lithium or with an AAP used in the model was based on data derived from a MTC carried out as part of this review (see *Chapter 3, Assessment of effectiveness*). A MTC was performed because no head-to-head studies were identified comparing SSRI augmentation with lithium and augmentation with a AAP. SSRI was used as the baseline treatment for the comparison. The studies included in the MTC used fluoxetine as the SSRI and olanzapine as the AAP, as no other studies were identified in a relevant population (see *Chapter 3, Assessment of effectiveness*). Consequently, there was

an absence of evidence to inform comparisons of different SSRI and AAP treatment strategies in a cost-effectiveness analysis. Therefore, based on the absence of evidence to suggest different efficacies associated with individual SSRIs or AAPs, a class effect was assumed for the SSRIs and the AAPs in the cost-effectiveness analyses. The robustness of this assumption was investigated in a sensitivity analysis that used data from a second MTC carried out as part of the clinical effectiveness analysis (see *Chapter 3, Assessment of effectiveness*). The second MTC included trials in which lithium, olanzapine, aripiprazole or quetiapine were used to augment a patient population treated with a mixture of SSRI or venlafaxine therapy. This sensitivity analysis does not explicitly relax the assumption of class effect (as no data were available). However, it does provide an estimate of the extent to which the cost-effectiveness results are sensitive to the use of additional data to inform of the class effect of AAP augmentation.

The probabilities of remission and response associated with each augmentation strategy were required to inform the decision tree component of the model, representing the acute phase of treatment. From the MTC, data were available for the outcome of response for both augmentation therapies. However, data on remission rate were available only for augmentation with an AAP. As a consequence of the limited data available to inform the acute treatment efficacy, the decision was taken to use an alternative approach to generate the required probabilities. The approach used involved sampling the treatment effect of each augmentation strategy (from a distribution of possible effects) and calculating the proportion of patients (in a cohort of 1000 for each treatment arm) that would achieve remission or response. The details of this alternative approach are described here and summarised in *Figure 15*.

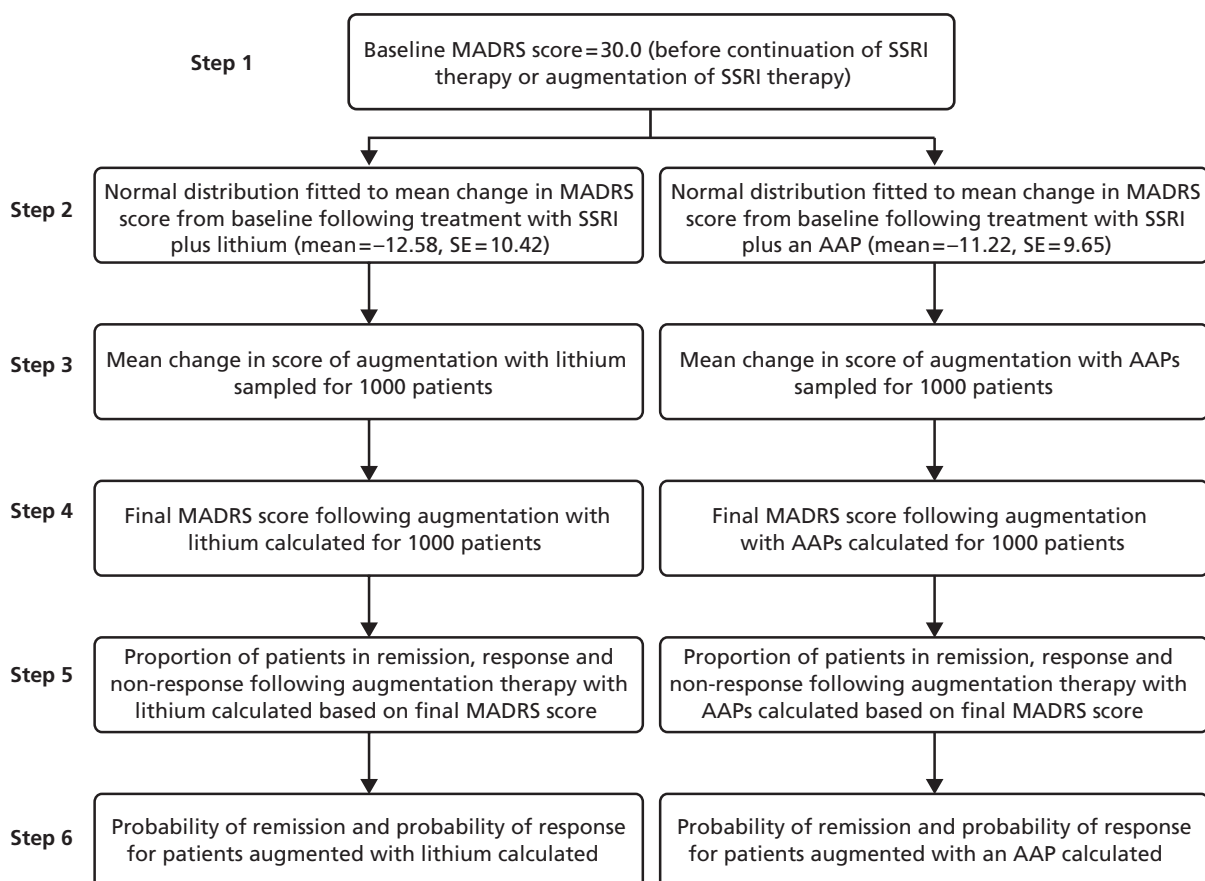


FIGURE 15 Calculating the treatment-specific probability of remission and response for the decision tree component of the model, representing the acute phase of treatment. SE, standard error.

Calculating the treatment-specific probability of remission and response

First, a baseline MADRS score of 30.0, which was considered to be representative of the TRD population (as defined in the scope of this review), was chosen and entered into the model. This baseline MADRS score was assumed to apply to all patients entering the model (i.e. before initiation of augmentation therapy). The average MADRS score of two studies pooled in an analysis by Thase *et al.*⁵³ identified as part of the clinical review was selected to provide the baseline MADRS score: the populations in these studies were considered to be the most representative of a TRD population compared with other studies identified as part of the clinical review (for more details of the clinical review see *Chapter 3, Assessment of effectiveness*).

Second, the absolute changes in MADRS score from baseline with SSRI + lithium and with SSRI + AAP were estimated from the MTC carried out as part of the clinical effectiveness analysis (see *Chapter 3, Assessment of effectiveness*). Lithium was associated with an absolute mean change in MADRS score of -12.58 [standard error (SE) 10.42] and the AAPs were associated with an absolute mean change in MADRS score of -11.22 (SE 9.65). The mean and SE of each treatment effect were used to fit a normal distribution from which the change in MADRS score associated with each augmentation therapy could be sampled. All patients in the cohort (1000 patients for each treatment arm) were then assigned a final MADRS score following augmentation of SSRI therapy with lithium or with an AAP (final MADRS score = baseline MADRS score + sampled change in MADRS score with augmentation treatment).

In clinical practice, the baseline MADRS score of the patient population would be variable. However, for the purposes of implementing the sampling methodology, the baseline MADRS score was assumed to be fixed, as the absolute treatment effect used to inform the sample distributions is not independent of baseline MADRS score. No data were available to account for the correlation between baseline MADRS score and the absolute effect of each augmentation treatment. Therefore, the decision was made to assume a fixed baseline MADRS score to avoid an inappropriate increase of the uncertainty associated with the acute treatment efficacy.

The final MADRS score of each patient following augmentation therapy was assessed against the definitions of response and remission outlined in *Chapter 3* (see *Assessment of effectiveness*). Therefore, patients who achieved a final MADRS score of < 10 were classified as remitters and patients who achieved a 50% or more reduction in their MADRS score from baseline (but did not achieve remission) were classified as responders. Patients who did not meet remission or the response criterion were classified as non-responders, defined as those patients not achieving a 50% reduction in MADRS score at the end of the acute treatment phase. The proportion of patients achieving remission, response and non-response for each augmentation therapy were calculated. These proportions were used to inform the probability of entering 'remission', 'response' and 'non-response' in the decision tree component of the model after 8 weeks of acute therapy.

Validation of acute efficacy estimation

The proportions of patients estimated to achieve response or remission by the sampling method described above were compared with the rates of response reported in the literature. Pooled analysis of two studies reported by Thase *et al.*⁵³ (selected to provide the baseline MADRS score used in the model) and the single lithium trial⁵⁹ identified for inclusion in the MTC were selected to provide this comparison. *Table 10* summarises the proportion of patients estimated (based on 2000 runs of 1000 samples for each treatment arm) to achieve remission and response and those reported in the literature.

The model estimates of response with AAP augmentation therapy and lithium augmentation therapy are approximately 5% lower than those reported in the literature.^{53,59} However, the baseline MADRS score of patients augmented with lithium in Katona *et al.*⁵⁹ is 26.06 (SD 4.93), rather than 30.0, as assumed in the model. In addition, the patient population of Katona *et al.*⁵⁹ had failed only one antidepressant treatment in their current episode of depression (for full details see *Chapter 3, Assessment of effectiveness*) and may be less resistant to treatment than the modelled patient population. Therefore, the difference in response rate between the model and Katona *et al.*⁵⁹ may be as a result of the difference in baseline depression severity. Furthermore, the efficacy estimate for augmentation with an AAP, used to inform the model, is based

TABLE 10 Comparison between response and remission probabilities estimated in the model and those reported in the literature

Outcome	Treatment	Model estimate (%)	Trial estimate (%)	Difference (model estimate – trial estimate): (%)	Trial reference
Response ^a	Lithium	43.0	47.1	–4.1	Katona <i>et al.</i> (1995) ⁵⁹
	AAPs	35.0	40.4	–5.4	Thase <i>et al.</i> (2007) ⁵³
Remission	Lithium	24.3	NR	N/A	N/A
	AAPs	18.7	27.3	–8.65	Thase <i>et al.</i> (2007) ⁵³
Non-response ^b	Lithium	57.1	N/A	N/A	N/A
	AAPs	65.0	N/A	N/A	N/A

N/A, not applicable; NR, not reported.

a Including remission.

b Probability of non-response was calculated as 1 – (the probability of response plus probability of remission).

on a MTC which included data from the two studies by Thase *et al.*⁵³ However, it is important to note that MD in MADRS score reported in Thase *et al.*⁵³ was much larger than that reported in the other trials included in the MTC. Therefore, estimates of response derived from the MTC may be expected to be lower than those reported in Thase *et al.*⁵³ The difference may be due to sampling error or heterogeneity in the studies included in the MTC.

Acute therapy discontinuation rates

Discontinuation may occur for a variety of reasons, although patients typically discontinue their acute treatment because of lack of efficacy or drug-related side effects. However, as discussed above (see *Model structure*), the model does not distinguish between reasons for discontinuation, and discontinuation is assumed to occur at week 4 of acute therapy. Rates of discontinuation (for any reason) were taken from the same MTC that provided acute efficacy estimates, carried out as part of the clinical efficacy analyses described in *Chapter 3* (see *Assessment of effectiveness*). Augmentation with an AAP was associated with a higher rate of discontinuation compared with no augmentation (OR 1.27; 95% CrI 0.90 to 1.75). However, augmentation with lithium was associated with a lower rate of discontinuation compared with no augmentation (OR 0.92; 95% CrI 0.13 to 3.32), albeit with wider 95% CrIs. These estimates were based on data from five AAP trials assessing a total of 1017 patients and a single lithium trial in 33 patients (reported in *Chapter 3*). As a consequence of the limited data available, the estimate of the OR of discontinuation with lithium augmentation is likely to be less reliable than that of AAP augmentation. This is demonstrated by the wider 95% CrIs associated with lithium augmentation.

To enable the use of the ORs estimated from the MTC in the economic model, the ORs were converted into relative risks (RRs) using the following formula:

$$RR = \frac{OR}{(1-r_b) + (r_b \times OR)} \quad (1)$$

where r_b = baseline risk of discontinuation (i.e. that associated with SSRI therapy alone) reported in the pooled analysis by Thase *et al.*⁵³

Clinical improvement following discontinuation from acute therapy

The model assumed that patients discontinuing their acute treatment at week 4 would not receive the benefit of treatment experienced by patients who complete 8 weeks of therapy. However, no studies were identified that reported response and remission rates for patients who did not complete their acute

treatment. Therefore, clinical experts were consulted to estimate, based on their experience, the likelihood of clinical improvement in patients who have discontinued from therapy. Experts agreed that overall approximately 5% and 15% of non-completers would remit and respond, respectively. These estimates were assumed to represent the annual probability of clinical improvement and standard formulae¹²¹ were used to convert them into 2-monthly probabilities for use in the model.

$$r = -[\ln(1-p)]/t \quad (2)$$

$$p_{2\text{-month}} = 1 - \exp\left(-\frac{r}{6}\right) \quad (3)$$

where p = annual probability, r = instantaneous rate and $p_{2\text{-month}}$ = 2-month probability.

Table 11 summarises the probabilities used to inform the decision tree component of the model.

Maintenance phase

The maintenance phase of the model includes 6 months of maintenance treatment and 4 months of follow-up (where patients may remain on or cease therapy depending on the level of response achieved). An absence of comparative long-term data for maintenance augmentation treatment led to the decision to assume that the maintenance phase of the model is predominantly treatment independent. Therefore, the parameters used in the Markov component of the model, representing the maintenance phase, are non-drug specific.

Table 12 summarises the probability associated with each possible transition for patients in the Markov model.

Response to remission

A longitudinal study conducted in the UK by Fekadu *et al.*¹²² followed 118 patients with TRD who had been discharged from specialist inpatient care (median follow-up of 3 years). Patient outcomes were reported by post-treatment status at discharge: remission, partial remission or still in episode. The authors reported that patients had received individualised treatment packages consisting of mainly pharmacotherapy

TABLE 11 All parameters used in the decision tree component

Parameter	Treatment-specific probability (%)		Sources
	Lithium	AAP	
Remission ^a	24.3	18.7	Sampling based on results of MTC
Response ^a	18.7	16.3	
Non-response ^a	57.0	65.0	
Discontinuation ^b	18.1	23.4	Thase <i>et al.</i> (2007) ⁵³ MTC
Remission following discontinuation	3	3	Expert clinical opinion
Response following discontinuation	1	1	
Non-response following discontinuation	96	96	

a Calculated from the PSA (1000 patients sampled 2000 times and the proportions of remission, response and non-response averaged).

b The rate of discontinuation associated with SSRI therapy reported in Thase *et al.*⁵³ is adjusted by the RRs of 0.93 and 1.21 for lithium and AAPs, respectively, estimated from the MTC.

TABLE 12 Transition probabilities (per 2-month cycle) used in the Markov model

Transition from	Transition to					
	Remission (%)	Remission discontinue (%)	Response (%)	Response discontinue (%)	Non-response (%)	Relapse (%)
Remission	70.3 ^a	20.8 ^b	0 ^c	0 ^c	0 ^c	8.9 ^d
Response	3.8 ^e	0 ^c	54.3 ^a	20.8 ^b	0 ^c	21.2 ^d
Remission discontinue	0 ^c	91.1 ^a	0 ^c	0 ^c	0 ^c	8.9 ^d
Response discontinue	0 ^c	3.8 ^d	0 ^c	75.1 ^a	0 ^c	21.2 ^d
Non-response	0 ^c	0.9 ^f	0 ^c	2.7 ^f	96.4 ^a	0 ^c
Relapse	0 ^c	0 ^c	0 ^c	0 ^c	0 ^c	1 ^c

a Figure = 1 – sum of all other transition probabilities from this health state.

b Assumed to be equal to the average of acute therapy discontinuation.

c Assumption.

d STAR*D trial.⁹⁹

e Fekadu *et al.*¹²²

f Expert opinion.

(mostly using medication combination), physical therapy and psychological therapy as indicated. The outcomes were assessed using the seven-point-scale Longitudinal Interval Follow-up Evaluation (LIFE) chart, which is a follow-up evaluation scale that allows retrospective rating of a patient's symptomatic state.¹²² The scores range from an asymptomatic state (score 1) to severe episode (score 7). Remission was defined as a score of 1 or 2, a score of 3 or 4 represented partial remission, and scores 5–7 were defined as a depressive episode. The authors reported that for those that were discharged in partial remission, 50% achieved remission over the 3-year follow-up period. This 3-year probability was converted into a 2-month probability using standard formulae.¹²¹ This probability was then used to represent the probability of improvement to 'remission' for patients in 'response' and for improvement to 'remission discontinuation' for patients in 'response discontinuation'. As acknowledged by the authors, the generalisability of the findings from Fekadu *et al.*¹²² may be limited because the patients included in the study received intensive specialist inpatient care at a specialist tertiary centre. Furthermore, patients who remain on therapy may be more likely to improve to the level required for remission than patients who have discontinued. However, in the absence of data to inform these different transitions in an appropriate population, data from Fekadu *et al.*¹²² were assumed to be applicable to the probability of remitting in the modelled population regardless of treatment status. This assumption is tested in the one-way sensitivity analysis (OWSA) (see *Results*).

Relapse

Relapse rates were based on the 1-year follow-up results reported for level 3 patients in the STAR*D trial.⁹⁹ The study evaluated feasible treatment strategies to improve clinical outcomes for real-world patients with TRD. Patients in level 3 had failed to benefit from at least two antidepressant treatment exposures. Relapse rates in this follow-up study were reported by treatment status, i.e. whether or not the patients had remitted prior to follow-up. Fewer patients who were in remission at the beginning of the follow-up period relapsed than those who were not in remission (42.9% vs. 76%). These annual probabilities were converted using standard formulae¹²¹ into 2-month probabilities for use in the model. The probability of relapse was assumed to be the same regardless of treatment status. Therefore, patients in 'response discontinuation' and 'remission discontinuation' were assigned the same probability of relapse as patients in 'response' and 'remission', respectively. This assumption is tested in the OWSA (see *Results*).

Non-response to remission discontinuation or response discontinuation

No study was identified to inform the probability of moving from a state of non-response to response or remission. However, expert clinical opinion sought during the construction of the acute treatment phase of the model was that approximately 5% and 15% of non-completers would remit and respond, respectively (see *Acute treatment phase*, above). Lack of efficacy is acknowledged to be a main reason for discontinuation. Therefore, it was considered reasonable to assume that the likelihood of entering response or remission following discontinuation from acute therapy is equivalent to the likelihood of remitting or responding for patients in non-response. The estimates obtained from expert clinical opinion (5% and 15% for remission and response, respectively) were assumed to represent the annual probability of remission or response and were converted into 2-monthly probabilities using standard formulae.¹²¹

Discontinuation

No studies were identified reporting discontinuation rates of patients on maintenance augmentation therapy. Therefore, discontinuation rates obtained for lithium augmentation and augmentation with an AAP from the MTC (18.1% and 23.4%, respectively) were averaged and assumed to apply in the maintenance phase, which resulted in a discontinuation rate of 20.8%.

Costs

The model is constructed from the perspective of the NHS and considers the costs of drug treatment, HCPs (e.g. GPs, and CMHTs), hospitalisation and treatment-related monitoring (e.g. laboratory tests). Social care costs, such as those related to residential care, were not included in the analysis because these costs were believed not to vary widely among the interventions assessed. Other societal costs, including social benefit payments, costs associated with legal system services and productivity losses of patients and carers were not estimated as they were beyond the scope of the analysis.

Drug costs

As discussed above (see *Acute treatment phase*), a class effect was assumed for the SSRIs and AAPs. Therefore, the cost of SSRI used in the model is a weighted average of the commonly prescribed SSRIs. The weights are based on clinical opinion rather than prescription cost analysis, as prescription cost analysis encompasses depression of any severity rather than focusing on TRD. The four most commonly prescribed SSRIs are citalopram (20%), escitalopram (20%), fluoxetine (30%) and sertraline (30%). However, clinical experts acknowledge that citalopram use is decreasing because of concerns over side effects of QT interval prolongation, particularly with respect to augmentation. The most commonly prescribed AAPs are aripiprazole, olanzapine, quetiapine and risperidone; experts estimated that the proportion of use for each of these would be 30%, 30%, 20% and 20%, respectively. Analogous to the assumptions made for SSRIs, a weighted average cost was calculated for AAPs. All drug costs were taken from the BNF 63.¹⁹ *Table 13* summarises the weighted drug costs used in the model. The maximum licensed dose of each SSRI is used in the cost calculations, as it is assumed that patients eligible for augmentation will be receiving the maximum dose of their current SSRI therapy. The dose assumed in the AAP cost calculation is the usual maintenance dose stated in the BNF.¹⁹ Lithium was assumed to be given in tablet form as the cost of tablets was lower than that of lithium injections (Priadel®, 800 mg per day). A threshold analysis around the cost of AAPs is carried out as part of the sensitivity analysis (see *Sensitivity analysis*, below). The cost of fluoxetine was also varied in sensitivity analysis using 3 × 20-mg tablets, which cost £4 per month, as the base case assumed the use of 1 × 60-mg fluoxetine tablet, which cost £53 per month.

Health-care professional costs

For the purpose of costing the patient pathway, the analysis assumes that during the acute phase of the disease a proportion of patients will be managed in the community and the remainder will be managed in hospital. For those patients managed in the community, a fraction will be seen by CRHTTs, who provide intensive home-based support, whereas the remainder will receive usual care from their GP and CMHT. Clinical experts estimate that 70% of patients with TRD will receive usual care and 20% will be seen by

TABLE 13 Commonly prescribed drugs and the weighted costs used in the model (SSRIs, AAPs and lithium)

Drug	No. in pack	Cost/pack (£)	Cost/day (£)	Cost/month (£)	Weighted drug use (%)	Weighted costs used in the model (£)	
						Acute	Maintenance
SSRIs							
Citalopram 40 mg	28	1.37	0.05	1	20	0.60	1.79
Sertraline 100 mg	28	1.80	0.13	4	30	2.35	7.04
Fluoxetine 60 mg	30	52.54	1.75	53	30	31.96	95.89
Escitalopram 10 mg	28	25.20	0.90	27	20	10.95	32.85
Total cost (£)						45.85	137.56
AAPs							
Quetiapine 300 mg	60	170.00	2.83	86	30	51.71	155.13
Olanzapine 5 mg	28	43.70	1.56	47	20	18.99	56.97
Aripiprazole 10 mg	28	95.74	3.42	104	30	62.40	187.21
Risperidone 3 mg	60	2.71	0.05	1	20	0.55	1.65
Total cost (£)						133.65	400.95
Lithium							
Priadel® 400 mg	100	3.35	0.07	2	100	4.08	12.23

the CRHHT. The remaining 10% of patients would receive inpatient care. These estimates are in line with those cited in the NICE Depression Guideline,¹⁶ as well as the King's Fund Report.²² Following acute treatment, all patients are assumed to receive care in the community, with the amount of care received depending on the level of response achieved.

Care in the community

Usual care

Usual care in this model refers to a package of care administered by the GP and the CMHT; patients receiving usual care are assumed to have no contact with a specialist psychiatrist. During the acute phase of treatment, it is assumed that patients receiving usual care will be seen by their GP and CMHT twice in the first month and once in the second month. During the maintenance phase of the model, patients in remission or response are assumed to see their GP and CMHT less frequently. Based on expert clinical opinion, patients who remit are assumed to have contact with their GP and CMHT five and three times, respectively, over the whole maintenance phase. Patients who respond but do not achieve remission are assumed to have five visits with their GP and CMHT over the whole maintenance phase.

Crisis Resolution and Home Treatment Teams

Crisis Resolution and Home Treatment Teams will visit about 20% of patients with TRD. CRHTTs work with patients in their own homes who, without this support, would need to be admitted to hospital. This ensures that 'patients recover at home in the environment in which they will have to function in the community rather than in an inpatient environment that bears little resemblance to normal living conditions'.¹²³ Following a crisis assessment, a care plan is agreed with the individual patient, which aims to meet their current mental health needs. For the purpose of this model, clinical experts suggested that during

the acute treatment phase patients will have contact with a CRHTT 16 times, i.e. he/she will be seen twice a week during the acute treatment phase. Following the acute treatment phase, patients who have not responded to treatment are assumed to be cared for by a CRHTT (one visit every 2 months plus one visit every 2 months to their GP). This assumption was made to represent the increased level (and cost) of care that is likely to be required by patients who do not respond to treatment. In addition, patients who relapse during the maintenance phase are assumed to be cared for by a CRHTT (one visit every 2 months plus one visit every 2 months to their GP).

The costs of care by a CMHT and a CRHTT were taken from the NHS reference costs for 2010–11,¹²⁴ and the cost of a visit to the GP was taken from the Unit Costs for Health and Social Care 2011.¹²⁵ All community care costs are summarised in *Table 14*.

Inpatient care costs

The model assumed that 10% of patients with TRD would be managed as inpatients. Data from the Hospital Episode Statistics (HES)¹²⁶ showed that the mean length of hospitalisation during acute depression for the year 2010–11 [F30–F39 Mood (affective) disorders] was 40 days (median 19 days). Therefore, the model assumed an inpatient length of stay of 40 days. The cost of inpatient mental health care was taken from the NHS reference costs for 2010–11.¹²⁴ The mean inpatient cost per day was calculated as a weighted average of costs for the following individual components: adult intensive care, acute care, rehabilitation and care of the elderly (codes MHIPA1–3 and MHIFE1). As summarised in *Table 15*, the cost of each component was weighted by the reported level of activity for that component to calculate the unit cost of inpatient care. After discharge from hospital, patients were assumed to receive usual care in the community by their GP and CMHT.

Monitoring costs

During initiation and maintenance of treatment, patients incurred the laboratory costs of blood testing required for monitoring purposes, depending on the type of long-term medication they received. The type and frequency of laboratory tests included in the model were provided by two clinical experts and checked for consistency with published clinical guidelines.^{16,123}

TABLE 14 Community care costs used in the model

HCPs	Average unit cost per visit used in the model (£)	Lower quartile unit cost (£)	Upper quartile unit cost (£)
CRHTT	179.00	136	207
CMHT	136.00	109	160
GP	36.00	30	40

TABLE 15 Inpatient costs used in the model

Currency description (NHS reference cost code)	Activity	National average costs (£)	Weight	Weighted mean costs (£)	Weighted lower quartile (£)	Weighted upper quartile (£)
Adult: Intensive Care (MHIPA1)	201,557	613	0.031566	19	16	21
Adult: Acute Care (MHIPA2)	2,994,811	304	0.469019	143	131	150
Adult: Rehabilitation (MHIPA3)	886,705	274	0.138867	38	31	44
Elderly (MHIFE1)	2,156,019	310	0.337656	105	94	116
Total weighted costs (£)				312	278	339

Patients receiving lithium have their lithium plasma levels monitored twice a month during the acute treatment phase then at 3-monthly intervals during maintenance treatment. Renal function tests are performed once at initiation and subsequently at 3-monthly intervals for the duration of treatment. Thyroid function tests are carried out once at initiation and biannually thereafter. Patients will also have electrocardiography, body mass index (BMI) and weight measured, and a full blood count at least once a year.

Patients receiving an AAP have a full blood count, urea and electrolyte levels measured, and a liver function test once a year. Fasting blood glucose and lipids profile are measured once at initiation and biannually thereafter. Weight should be measured at least once a month in the acute phase and then at 3-monthly intervals for the remaining duration of treatment. The costs of laboratory tests and monitoring were taken from the NHS reference costs and literature.¹²⁴ Tables 16 and 17 summarise the frequency and type of test carried out during the acute and maintenance phases of treatment for those patients receiving augmentation with lithium or an AAP, respectively.

Costs associated with each health state

Acute treatment phase

The medication cost incurred by patients in the acute treatment phase depends on whether a patient completes their acute therapy or discontinues. As discussed in *Model structure*, above, patients who discontinue their acute treatment are assumed to do so after 4 weeks of therapy. Therefore, those patients who do discontinue will incur the cost of 4 weeks rather than 8 weeks of acute treatment. However,

TABLE 16 Frequency of tests and total costs for patients augmented with lithium during the acute and maintenance phases

Laboratory costs	Unit cost (£)	Source	Acute treatment phase	Maintenance treatment phase	Follow-up phase (if treatment is continued)	Comment
BMI	6.17	Meads <i>et al.</i> (2008) ¹²⁷	1	0	0	Once a year
ECG	34.94	NHS ref costs 2010–11 ¹²⁴	1	0	0	Once a year
Full blood count	2.35	^a Bipolar guidelines ¹²³	1	0	0	Once a year
eGFR	1.03	^a Bipolar guidelines ¹²³	1	2	1	Once during initiation, and then at 3-monthly intervals for the duration of treatment
Creatinine (to monitor renal function)	1.03	^a Bipolar guidelines ¹²³	1	2	1	Once during initiation, and then at 3-monthly intervals for the duration of treatment
Serum lithium concentration (lithium plasma levels)	2.82	^a Bipolar guidelines ¹²³	4	2	1	Once a month during initiation then at 3-monthly intervals for the duration of treatment
Thyroid function	16.08	^a Bipolar guidelines ¹²³	1	1	1	Once during initiation and then biannually
Cost applied in the model (£)			72.85	25.82	20.95	

eGFR, estimated glomerular filtration rate.

^a Inflated to 2011 prices using the Hospital and Community Health Services inflation indices – Curtis 2010.¹²⁵

TABLE 17 Frequency of tests and total costs for patients augmented with an AAP during the acute and maintenance phases

Laboratory costs	Unit cost (£)	Source	Acute treatment phase	Maintenance treatment phase	Follow-up phase (if treatment is continued)	Comment
BMI	6.17	Meads 2008 ¹²⁷	2	2	1	Once a month during initiation then at 3-monthly intervals for the duration of treatment
ECG	34.94	NHS reference costs 2010–11 ¹²⁴	1	0	0	Once a year
Full blood count	2.35	Bipolar guidelines ^a	1	0	0	Once a year
eGFR	1.03	Bipolar guidelines ^a	1	0	0	Once a year
Creatinine (to monitor renal function)	1.03	Bipolar guidelines ^a	1	0	0	Once a year
Glucose test	1.03	Bipolar guidelines ^a	1	1	1	Once during initiation and then biannually
Lipid profile test	2.21	Bipolar guidelines ^a	1	1	1	Once during initiation and then biannually
Cost applied in the model (£)			54.91	18.80	9.40	

eGFR, estimated glomerular filtration rate.

^a Inflated to 2011 prices using the Hospital and Community Health Services inflation indices, Curtis 2010.¹²⁵

monitoring costs in the acute phase are assumed to remain the same regardless of treatment status. Furthermore, all patients in the acute treatment phase will receive the same cost of acute patient care regardless of which health state they enter. As discussed in *Health-care professional costs*, above, some patients receive care in the community while others are cared for as inpatients. The cost of patient care in the acute treatment phase is calculated as a weighted average of community and inpatient care (*Table 18*). *Table 19* summarises the total cost applied to patients in the acute treatment phase.

Maintenance phase

The costs accrued in the maintenance phase of the model depend on a patient's health state. *Table 20* summarises the costs associated with each health state in the maintenance phase. As a simplification, the cost per cycle is calculated as a proportion of the total cost had the patient remained in state throughout the maintenance phase. For example, the cost per cycle of patients in remission is calculated as one-fifth of the total cost of entering the maintenance phase in remission and remaining in remission for the duration of the model. This assumption was considered reasonable, as the costs associated with HCPs, drugs and monitoring are likely to be evenly distributed across the 10 months of the maintenance phase.

Costs for patients who remit

Patients who remit are assumed to be under the care of HCPs for a total of 6 months in the maintenance phase. At the end of maintenance treatment, patients in remission are assumed to be discharged from the care of their CMHT and will subsequently be cared for by their GP alone. For the remaining 4 months of the model, following the completion of maintenance therapy, remitters are assumed to visit their GP once every 2 months for a general check-up. Therefore, the costs accrued by a patient in remission during

TABLE 18 Cost of patient care applied in the acute treatment phase (HCPs)

Setting of care	Proportion of patients (%)	Cost component	Cost (£)	Resource use during acute treatment phase	Total cost (£)
Community (CRHTT)	20	CRHTT visit	179	16 visits	$(16 \times 179) = 2864$
Community (usual care)	70	GP visit	36	2 visits	$(2 \times 36) + (2 \times 136) = 344$
		CMHT visit	136	2 visits	
Inpatient	10	Mental health inpatient cost	312	40 days	$(40 \times 311.96) + 136 + 36 = 12,650.40$
		GP visit	36	1 visit	
		CMHT visit	136	1 visit	
Overall weighted cost (£)					$(0.2 \times 2864) + (0.7 \times 344) + (0.1 \times 12,650.4) = 2078.64$

TABLE 19 Total costs applied to patients in the acute treatment phase

Treatment status	Cost component	Cost (£)		Total cost (£)	
		Lithium	AAP	Lithium	AAP
Completed acute therapy	Eight weeks of acute treatment	49.93	179.50	2201.42	2313.05
	Acute monitoring costs	72.85	54.91		
	Eight weeks of acute patient care	2078.64	2078.64		
Discontinued acute therapy	Four weeks of acute treatment	24.96	89.75	2176.46	2223.30
	Acute monitoring costs	72.85	54.91		
	Eight weeks of acute patient care	2078.64	2078.64		

the Markov component of the model would include 6 months of maintenance treatment; maintenance monitoring; three visits with the GP and a CMHT during maintenance treatment (one visit every 2 months); and two further visits to his/her GP.

Costs for patients who respond

In the Markov component of the model, responding patients are assumed to remain on maintenance therapy for the full model time horizon. In addition, responding patients are assumed to continue to receive usual care by their GP and CMHT. Therefore, responders incur 10 months of maintenance treatment, monitoring for the maintenance and follow-up phase, and five visits with their GP and CMHT (one visit every 2 months).

Costs of non-responding patients

Patients who do not respond to acute treatment enter the non-response health state. These patients are taken off their augmentation therapy and are assigned a standard package of care that includes:

- 10 months of non-specific therapy (calculated by averaging the cost of augmentation with lithium and with an AAP)
- patient care provided by a CRHTT and GP (one visit every 2 months by CRHTT and GP)
- monitoring costs (averaged of lithium and AAP monitoring costs).

TABLE 20 Cost per cycle associated with each health state in the maintenance phase by treatment arm

Health state	Cost components	Cost per cycle (£) over the 10 months of the Markov model	
		Lithium	AAP
Remission	Six months of maintenance treatment Three CMHT visits during maintenance treatment phase Five GP visits (three during maintenance treatment phase and two during follow-up) Maintenance monitoring	127.27	190.89
Response	Ten months of maintenance treatment Five visits with a CMHT and GP (three during the maintenance treatment phase and two during follow-up) Maintenance and follow-up monitoring	192.74	297.62
Relapse	Standard package of care: Ten months of non-specific therapy ^a Patient care provided by a CRHTT and GP (one visit every 2 months) Monitoring costs (maintenance and follow-up)	281.01	281.01
Discontinuation remission	Three CMHT visits during maintenance treatment phase Five GP visits (three during maintenance treatment phase and two during follow-up) Two GP visits during follow-up	98.00	98.00
Discontinuation response	Five visits with a CMHT and GP (three during the maintenance treatment phase and two during follow-up)	143.33	143.33
Non-response	Standard package of care: <ul style="list-style-type: none">• Ten months of non-specific therapy^a• Patient care provided by a CRHTT and GP (one visit every 2 months)• Monitoring costs (maintenance and follow-up)	281.01	281.01

^a Cost assumed to be equal to the average cost of augmentation therapy.

Costs for patients who relapse

A patient in the fully symptomatic health state (MADRS score of > 15) after having achieved remission or response is considered to be in the relapse state. Patients who relapse during the maintenance phase of the model (either during maintenance treatment or in the 4 months following maintenance treatment) are assumed to receive the standard package of care received by patients in 'non-response' at the end of the acute treatment phase for the remainder of the model.

Costs for patients who discontinue following remission or response

Patients who discontinue maintenance therapy after experiencing remission or response are assumed to receive the same level of patient care as their counterparts who remain on therapy. Therefore, patients who discontinue following remission would incur the cost of 6 months of usual care (GP plus CMHT) and two further visits to their GP. Patients who discontinue following response to treatment would accrue the cost of 10 months of usual care.

Sensitivity analysis

Probabilistic sensitivity analysis was the standard method of analysis, and mean costs and QALYs from the PSA were used to inform the base-case results. In addition, several sensitivity analyses were carried out in support of the economic evaluation of augmentation of SSRI therapy with lithium compared with an AAP. These were:

- OWSA of cost, acute efficacy and acute discontinuation, to assess the univariate sensitivity of the deterministic model to changes in individual parameters
- a threshold analysis on the cost of AAPs
- scenario analysis, to assess the sensitivity of the model to the assumption of class effects.

Probabilistic sensitivity analyses

To assess the simultaneous effect of parameter uncertainty on outcomes of cost-effectiveness, the model was built to be fully probabilistic. This was achieved by assigning appropriate distributions to each parameter from which estimated values were repeatedly sampled. Samples of 1000, 2000 and 5000 were used in the base-case model and the stability of the model was assessed with respect to non-linearity between the deterministic and probabilistic results. A sample size of 2000 was chosen, as this gave a more stable probabilistic estimate than a sample size of 1000 and was more efficient than a sample size of 5000. *Table 21* displays the distributions assigned to each model parameter.

Uncertainty in the acute treatment phase

Uncertainty around the effect of acute augmentation treatment was captured by assigning a normal distribution to the absolute treatment effect estimated from the MTC carried out as part of the clinical effectiveness analysis (see *Chapter 3, Assessment of effectiveness*). However, uncertainty around patients' baseline MADRS scores has not been captured, as explained above (see *Acute treatment phase*). This is because it was considered that the potential correlation between baseline MADRS score and treatment effect would compromise the assumption of parameter independence necessary for a PSA unadjusted for correlation as described above (see *Acute treatment phase*).

The probability of patients responding or remitting following discontinuation of acute treatment was informed by expert clinical opinion. Therefore, the uncertainty associated with these estimates was not readily quantifiable in terms of a SE or CI. Consequently, the uncertainty around these estimates was approximated by examining the effect of varying the estimates within a boundary of $\pm 25\%$ using a log-normal distribution.

Discontinuation in the acute treatment phase was informed by the MTC carried out as part of the clinical effectiveness review (see *Chapter 3, Assessment of effectiveness*). Uncertainty around these estimates was captured by assigning a log-normal distribution to the ORs obtained from the MTC.

Uncertainty in the maintenance phase

The maintenance phase of the model, which includes 6 months of maintenance treatment plus 4 months of follow-up (where patients may remain on or cease treatment, depending on the level of response achieved) is parameterised by the following probabilities:

- relapse for patients currently in 'remission' (or 'remission discontinuation')
- relapse for patients currently in 'response' (or 'response discontinuation')
- remission for patients currently in 'response' discontinuation
- remission for patients currently in 'non-response' (these patients will enter the health state of 'remission discontinuation')
- response for patients currently in 'non-response' (these patients will enter the health state of 'remission discontinuation').

TABLE 21 Summary of parameters and distributions implemented in the PSA

Parameter	Mean value	SE/ variation	Distribution used
Acute efficacy			
Acute efficacy of lithium augmentation therapy (change in MADRS score)	-12.58	10.42	Normal
Acute efficacy AAP augmentation therapy (change in MADRS score)	-11.22	9.65	Normal
Probability of clinical improvement for non-completers/non-responders			
Annual probability of remission	5%	± 25%	Log-normal
Annual probability of response	15%	± 25%	Log-normal
Acute discontinuation			
OR for discontinuation in patients treated with lithium + SSRI vs. SSRI alone	0.92	0.85	Log-normal
OR for discontinuation in patients augmented with an AAP + SSRI vs. SSRI alone	1.27	0.17	Log-normal
Maintenance-phase parameters			
Three-year probability of moving from response to remission	0.500	0.064	Beta
Annual probability of relapse for patients who have responded to treatment	0.760	0.052	Beta
Annual probability of relapse for patients who are in remission	0.439	0.082	Beta
Patient care costs (£)			
CRHTTs	179.00	179.00	Gamma
CMHTs	136.00	136.00	Gamma
Mental health inpatients	318.00	318.00	Gamma
Outpatient setting: first attendance	228.00	228.00	Gamma
Outpatient setting: follow-up	160.00	160.00	Gamma
Mental Health Secure Units	533.45	533.45	Gamma
GP	36.00	36.00	Gamma
Monitoring costs (£)			
BMI	6.17	6.17	Gamma
ECG	34.94	34.94	Gamma
Full blood count	2.35	2.35	Gamma
eGFR	1.03	1.03	Gamma
Creatinine	1.03	1.03	Gamma
Serum lithium concentration	2.82	2.82	Gamma
Thyroid function	16.08	16.08	Gamma
Glucose test (Biochemistry EAP 841)	1.03	1.03	Gamma
Lipid profile test	2.21	2.21	Gamma
Utility values			
Remitters	0.8500	0.008	Beta
Responders	0.7200	0.013	Beta
Non-responders ^a	0.5800	0.018	Beta

eGFR, estimated glomerular filtration rate.

^a Assumed to be the same at baseline as patients with severe depression.

Beta distributions were assigned to the probabilities derived from literature sources, and the SEs associated with these probabilities were calculated from the data reported in the relevant literature source. Therefore, the uncertainty around the probability of remission (for patients currently in 'response' or 'response discontinuation') and of relapse was calculated from data provided in the studies of Fekadu *et al.*¹²² and Rush *et al.*,⁹⁹ respectively. The log-normal distribution was used to assess the uncertainty around probabilities estimated from expert clinical opinion, with arbitrary variation of $\pm 25\%$ assumed. This is because the sample size of the expert panel was considered to be too small ($n = 2$) to calculate the alpha and beta values required to implement the beta distribution. Discontinuation in the maintenance phase of the model is assumed to be an average of the treatment-specific discontinuation rates of the acute treatment phase. Therefore, uncertainty around discontinuation in the maintenance phase is accounted for by assessing the uncertainty around discontinuation in the acute treatment phase.

Uncertainty around cost and quality of life

It is assumed that there is no uncertainty surrounding the current cost of drugs to the NHS; therefore, drug costs were not varied in the sensitivity analysis. The impact of uncertainty surrounding the costs of monitoring and patient care was assessed by assigning gamma distributions to these costs (see *Table 21*). The gamma distribution was chosen as cost data are often highly skewed and the gamma distribution is constrained to be non-negative.¹²¹ However, as SEs were not available (and could not be derived) for these costs, the SE of each cost was assumed to be equal to the mean cost.¹²¹ Uncertainty around the values used for utility was assessed using the beta distribution (see *Table 21*) based on SEs reported in the study by Sapin *et al.*¹⁰⁵

The analysis was run over 2000 iterations, and the expected costs and QALYs generated from each run were averaged to determine the probabilistic ICER reported below (see *Results*) for the base-case and scenario analyses.

One-way sensitivity analyses

A range of OWSA was undertaken on key model inputs to investigate the impact the independent variation of each input had on the incremental cost, incremental benefit and ICERs. Each key parameter was alternately assigned a low and high value and the deterministic cost-effectiveness results using this value recorded. The deterministic model was used for the OWSA as a pragmatic measure because the probabilistic model was computationally intensive (average run time ~ 10 hours).

In the OWSA, it was assumed that there is no uncertainty with regards to current drug costs in the NHS; therefore, drug costs were not varied in these analyses. The interquartile ranges reported in the National Schedule of Reference Costs were used as the low and high values of HCP costs. However, the variation associated with the cost of monitoring was not reported; therefore, the values used in the OWSA were assumed to be $\pm 25\%$ of the model's base-case value.

To assess the impact of univariate changes in acute treatment efficacy, 95% CIs were calculated from the means and SEs reported in the MTC (used to inform the normal distribution from which treatment effect is sampled in the base case) for each treatment. Then for each treatment, in turn, the upper and lower thresholds of the CI associated with that treatment were alternatively used to represent the mean change in MADRS score. The mean change in MADRS score was then used to inform the normal distribution from which acute treatment efficacy is sampled. The upper and lower values used in the assessment of the model's sensitivity to acute discontinuation were derived from the 95% CI associated with the log-normal distribution assigned to acute discontinuation (see discussion of PSA).

The uncertainty surrounding the assumption that the probabilities of relapse and of improvement (for patients in response) are independent of treatment status was assessed using assumed variance of $\pm 25\%$.

Table 22 summarises the upper and lower values used in each univariate sensitivity analysis. The results of this and other sensitivity analyses are presented below (see *Results*). In addition, any parameters identified as being significant in the OWSA based on the deterministic model were assessed in the probabilistic model.

TABLE 22 Upper and lower values used to inform univariate sensitivity analyses

Parameter	Base-case value	Lower value	Upper value	Source of variation
HCP costs (£)				
CRHTTs	179.00	136	207	Interquartile ranges reported in National Schedule of Reference Costs
CMHTs	136.00	109	160	
GP	36.00	30	40	
Mental health inpatient costs	312	278	339	
Monitoring costs (£)				
BMI	6.17	4.62	7.71	Assumed to be 25%
ECG	34.94	26.20	43.67	
Full blood count	2.35	1.76	2.94	
eGFR	1.03	0.77	1.28	
Creatinine	1.03	0.77	1.28	
Serum lithium concentration	2.82	2.11	3.52	
Thyroid function	16.08	12.06	20.10	
Glucose test (Biochemistry EAP 841)	1.03	0.77	1.28	
Lipid profile test	2.21	1.66	2.76	
Acute treatment efficacy				
Change in MADRS score following treatment with SSRI plus lithium	Sampled from normal distribution, mean -12.58	Sampled from normal distribution, mean -33.00	Sampled from normal distribution, mean 7.84	95% CI estimated from results of MTC
Change in MADRS score following treatment with SSRI plus an AAP	Sampled from normal distribution, mean -11.22	Sampled from normal distribution, mean -30.13	Sampled from normal distribution, mean 7.69	
Acute discontinuation				
Acute lithium discontinuation	OR 0.92	OR 0.13	OR 1.75	95% CI associated with log-normal distribution (fitted to ORs as part of PSA)
Acute AAP discontinuation	OR 1.27	OR 0.90	OR 3.32	
Maintenance-phase probabilities				
Remitting for patients in 'response discontinuation'	3.8%	2.9%	4.8%	Assumed to be ±25%
Relapse for patients in 'response discontinuation'	21.2%	15.9%	26.5%	Assumed to be ±25%
Relapse for patients in 'remission discontinuation'	8.9%	6.7%	11.1%	Assumed to be ±25%
eGFR, estimated glomerular filtration rate.				

Threshold analysis on the cost of atypical antipsychotic drugs

At present, there is significant uncertainty surrounding the future cost of AAPs to the NHS. This is because at the time of writing this report the patents held by Eli Lilly for olanzapine and AstraZeneca for quetiapine were close to expiration. (Note that by the time of publication of this report these patents had expired.) In addition, the patent for aripiprazole held by Otsuka will expire in 2015–16.¹²⁸ Therefore, the price of AAPs is likely to decrease as the manufacturers face competition from the generic market. However, there is uncertainty over the extent of the price reduction expected. A threshold analysis on the price of aripiprazole, olanzapine and quetiapine used in the base case was carried out to explore the impact of various price reductions on the base-case model results. The results of this are displayed below (see *Results*).

Scenario analyses

In the base case, the model assumes a class effect for the SSRIs and the AAPs. This assumption was based on an absence of evidence of a difference of effect, rather than evidence of no difference in effect. Therefore, there is a degree of uncertainty surrounding this assumption. Moreover, data were not available to relax this assumption. However, a scenario analysis was carried out to assess the sensitivity of the model to the use of additional data to inform the class effect of AAP augmentation. In the scenario analysis, data were used from a MTC carried out as part of the clinical effectiveness sensitivity analysis (see *Chapter 3, Assessment of effectiveness*). The MTC included trials in which lithium, olanzapine, aripiprazole or quetiapine were used to augment a patient population treated with a mixture of SSRI or venlafaxine therapy. *Table 23* summarises the acute efficacy and discontinuation derived from the MTC and used in this scenario analysis. The results of this analysis are presented below (see *Results*).

Results

Base-case results

The probabilistic and deterministic base-case results of the comparison between augmentation of SSRI therapy with lithium and augmentation with an AAP are displayed in *Table 24*. Deterministic results are reported to facilitate understanding on the OWSA results. The cost-effectiveness plane and cost-effectiveness acceptability curve (CEAC) associated with the probabilistic base case are displayed in *Figures 16* and *17*, respectively.

The base-case results indicate that augmentation of SSRI therapy with an AAP is dominated by augmentation of a SSRI with lithium. The cost-effectiveness plane (see *Figure 16*) represents the incremental costs and

TABLE 23 Acute efficacy and discontinuation used in the scenario analysis of class effect

Parameter	Treatment-specific probability (%)		Sources
	Lithium	AAP	
Remission ^a	16.7	20.7	Sampling based on results of MTC
Response ^a	17.3	16.4	
Non-response ^a	66.0	62.9	
Discontinuation ^b	18.0	23.0	Thase <i>et al.</i> (2007) ⁵³ MTC

a Calculated from the PSA (1000 patients sampled 2000 times and the proportions of remission, response and non-response averaged).

b The rate of discontinuation associated with SSRI therapy reported in Thase *et al.*⁵³ is adjusted by the RRs of 0.93 and 1.18 for lithium and AAPs, respectively, estimated from the MTC.

TABLE 24 Base-case probabilistic and deterministic results of the comparison between augmentation of SSRI therapy with lithium and augmentation with an AAP

Augmentation treatment	Cost (£)	QALYs	Incremental cost (£)	Incremental QALYs	ICER (£/QALY)
Probabilistic results^a					
Lithium	4739	1.253	–	–	–
AAP	5644	1.225	905	–0.03	AAP is dominated by lithium
Deterministic results					
Lithium	4702	1.258	–	–	–
AAP	5643	1.226	941	–0.03	AAP is dominated by lithium

a Probabilistic results were generated by averaging the results of 2000 probabilistic runs.

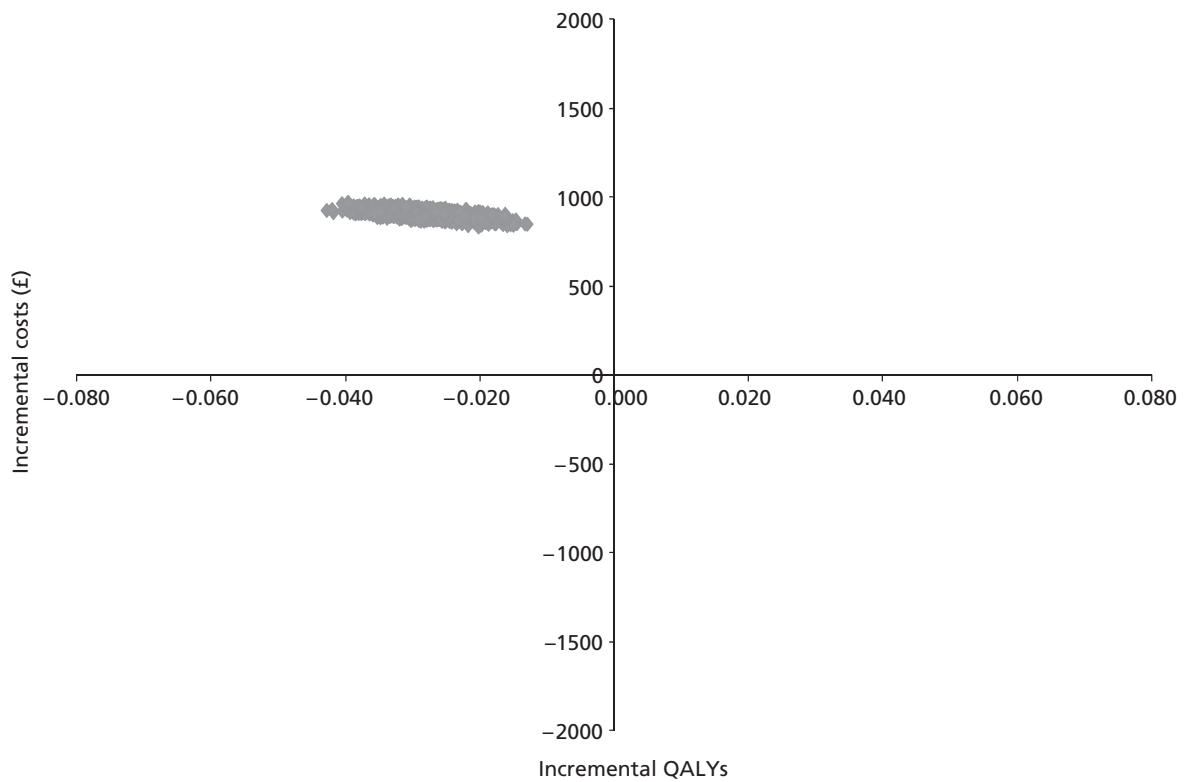


FIGURE 16 Cost-effectiveness plane associated with base-case probabilistic analysis of augmentation of SSRI therapy with lithium vs. augmentation with an AAP.

QALYs associated with AAP augmentation (vs. lithium augmentation) as points on the graph. This scatter plot suggests that there is no uncertainty about the dominance result and it is important to note that AAP augmentation did not provide more benefit than lithium augmentation in any of the probabilistic runs.

One-way sensitivity analysis

One-way sensitivity analysis of cost, acute efficacy and acute discontinuation were carried out on the deterministic base-case result, the results of which are presented in *Table 25* and the tornado diagram displayed in *Figure 18*.

The results of the OWSA indicated that the key drivers of the cost-effectiveness results are acute efficacy and discontinuation. Changes in assumptions around the probability of relapse and the probability of remission in

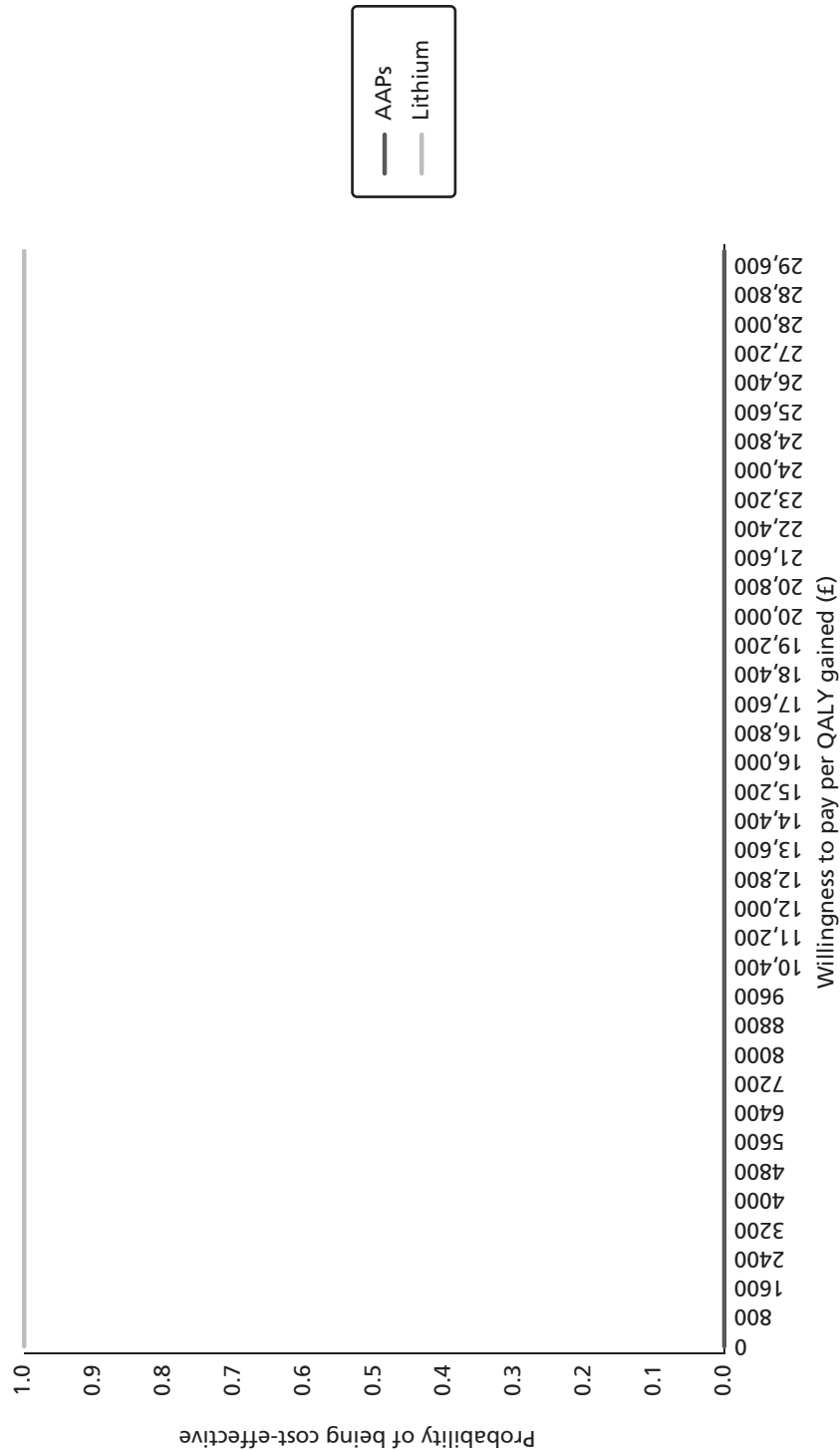


FIGURE 17 Cost-effectiveness acceptability curve associated with the probabilistic base case of augmentation of SSRI therapy with lithium vs. augmentation with an AAP.

TABLE 25 Results of OWSAs

Parameter	Incremental costs (£)	Incremental QALY	Cost (£)/QALY (ICER)
Base case	941	-0.03	Lithium dominates
Discontinuation, AAPs			
Lower 95% CI	964	-0.03	Lithium dominates
Upper 95% CI	917	-0.04	
Discontinuation, lithium			
Lower 95% CI	1079	-0.05	Lithium dominates
Upper 95% CI	700	0.01	113,596 ^a
Change in MADRS score, AAPs			
Lower 95% CI	917	0.20	4672 ^b
Upper 95% CI	815	-0.11	Lithium dominates
Change in MADRS score, lithium			
Lower 95% CI	2026	-0.26	Lithium dominates
Upper 95% CI	169	0.08	1996 ^b
CRHTT			
Lower costs	836	-0.03	Lithium dominates
Upper costs	992	-0.03	
CMHT			
Lower costs	941	-0.03	Lithium dominates
Upper costs	941	-0.03	
GP			
Lower costs	936	-0.03	Lithium dominates
Upper costs	941	-0.03	
BMI			
Lower costs	941	-0.03	Lithium dominates
Upper costs	944	-0.03	
ECG			
Lower costs	941	-0.03	Lithium dominates
Upper costs	943	-0.03	
Full blood count			
Lower costs	941	-0.03	Lithium dominates
Upper costs	941	-0.03	
Creatinine			
Lower costs	941	-0.03	Lithium dominates
Upper costs	941	-0.03	
Serum lithium concentration			
Lower costs	941	-0.03	Lithium dominates
Upper costs	938	-0.03	

TABLE 25 Results of OWSAs (continued)

Parameter	Incremental costs (£)	Incremental QALY	Cost (£)/QALY (ICER)
Thyroid function			
Lower costs	941	-0.03	Lithium dominates
Upper costs	937	-0.03	
Glucose test			
Lower costs	941	-0.03	Lithium dominates
Upper costs	942	-0.03	
Lipid profile test			
Lower costs	941	-0.03	Lithium dominates
Upper costs	942	-0.03	
eGFR			
Lower costs	941	-0.03	Lithium dominates
Upper costs	941	-0.03	
Inpatients costs			
Lower costs	941	-0.03	Lithium dominates
Upper costs	969	-0.03	
Response to remission rates in patients still on maintenance therapy (lower value)	941	-0.03	Lithium dominates
Response to remission rates in patients still on maintenance therapy (higher value)	941	-0.03	Lithium dominates
Relapse rates in response patients who have discontinued therapy (lower value)	942	-0.03	Lithium dominates
Relapse rates in response patients who have discontinued therapy (higher value)	940	-0.03	Lithium dominates
Relapse rates in remission patients who have discontinued therapy (lower value)	941	-0.03	Lithium dominates
Relapse rates in remission patients who have discontinued therapy (higher value)	941	-0.03	Lithium dominates

eGFR, estimated glomerular filtration rate.

a AAPs have an ICER of > £20,000/QALY.

b AAPs have an ICER of < £20,000/QALY.

the maintenance phase had minimal impact on the overall results. Similarly, changes in costs did not change the overall result of lithium dominance. However, changes in acute efficacy or discontinuation reversed the direction of the cost-effectiveness results. For example, when a low level of acute efficacy (upper 95% CI) for lithium was assumed, the direction of cost-effectiveness reversed and AAP augmentation had ICERs of < £20,000/QALY compared with lithium augmentation, suggesting that AAPs will be the preferred strategy. A similar result was observed when a high level of acute efficacy for the AAPs (lower 95% CI) was assumed. When higher levels of discontinuations were assumed for lithium, AAP augmentation resulted in more health benefits, albeit with an estimated ICER of > £100,000/QALY. Therefore, the impact of univariate changes in these key parameters (acute efficacy and discontinuation) on the results of the probabilistic model were examined. The results of these analyses are displayed in *Table 26* and the tornado diagram in *Figure 19*.

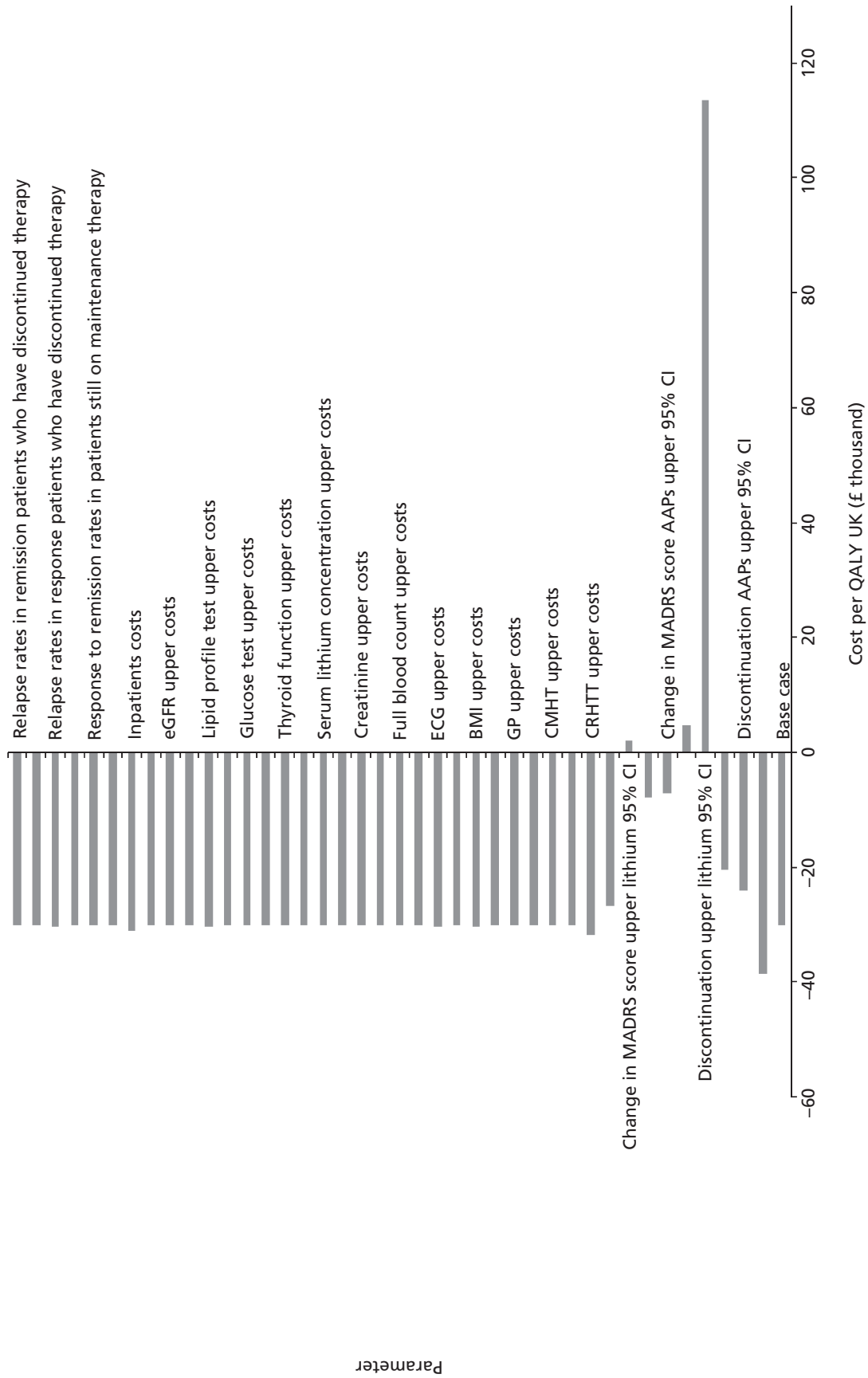
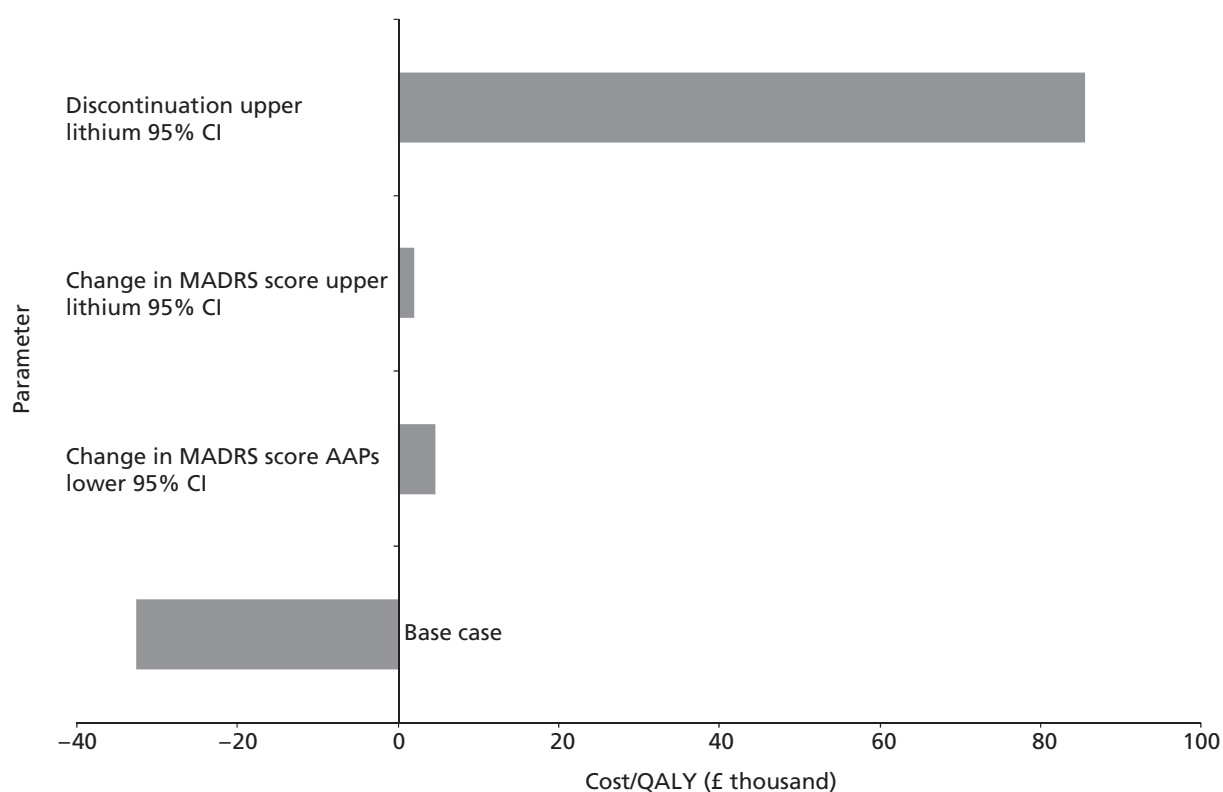


FIGURE 18 Results of the OWSAs on the deterministic base case of augmentation of SSRI therapy with lithium vs. augmentation with an AAP. eGFR, estimated glomerular filtration rate.

TABLE 26 Impact of OWSAs (response and discontinuation) on the results of the probabilistic model of augmentation of SSRI therapy with lithium vs. augmentation with an AAP

Parameter	Incremental costs (£)	Incremental QALY	Cost (£)/QALY (ICER)
Base case	905	-0.03	-32,650
Change in MADRS score (response)			
Change in MADRS score, AAPs lower 95% CI	916	0.20	4688
Change in MADRS score, upper lithium 95% CI	174	0.09	1976
Discontinuations			
Discontinuation, upper lithium 95% CI	676	0.01	85,600

**FIGURE 19** Impact of the OWSAs (response and discontinuation) on the results of the probabilistic model of augmentation of SSRI therapy with lithium vs. augmentation with an AAP.

The mean expected values from the probabilistic model compare well with the deterministic model, indicating that the results are robust to plausible changes in parameter estimates.

Threshold analysis

As discussed above (see *Sensitivity analysis*), a threshold analysis on the cost of AAPs to the NHS was carried out to assess the uncertainty around these costs, in the current climate of patent expiry. The results of this sensitivity analysis are presented in *Table 27*.

The results of this threshold analysis further highlight the robustness of the model results to changes in costs.

TABLE 27 Results of the threshold analysis of AAP drug cost

Parameter	Incremental costs (£)	Incremental QALY	Cost (£)/QALY (ICER)
Base case	941.06	-0.03	-30,206
AAPs fall in price (%)			
60	786.47	-0.03	-25,244
80	734.94	-0.03	-23,590
90	709.18	-0.03	-22,763
Fluoxetine (non-proprietary) 20 mg	919	-0.03	-29,512

Scenario analyses

As discussed above (see *Sensitivity analyses*), a scenario analysis was carried out in support of the economic evaluation of augmentation of SSRI therapy with lithium compared with an AAP. In this scenario analysis, additional data from a sensitivity analysis carried out as part of the clinical effectiveness section were used to inform the class effect of AAP augmentation. The clinical sensitivity analysis was a MTC including trials in which lithium, olanzapine, aripiprazole or quetiapine was used to augment a patient population treated with a mixture of SSRI or venlafaxine therapy. The probabilistic and deterministic results of this are displayed in *Table 28*, with the cost-effectiveness plane and CEAC associated with the probabilistic result displayed in *Figures 20* and *21*, respectively.

Compared with the base case, the scenario analysis into class effect results in a reduction in the incremental benefit of augmentation therapy with lithium.

Health economics discussion

The base-case results indicated that augmentation of SSRI therapy with an AAP is dominated by augmentation of a SSRI with lithium in patients with TRD. Furthermore, PSA suggested that the benefit obtained with AAP augmentation was consistently dominated by the benefit obtained with lithium augmentation; none of the probabilistic runs resulted in an incremental QALY gain for AAP augmentation over lithium augmentation. However, OWSA revealed that the model is highly sensitive to the relative level of discontinuation assumed and changes in the distributions used to sample the acute efficacy of treatment.

TABLE 28 Probabilistic and deterministic results of the scenario analysis extending the assumption of class effect based on data from a clinical sensitivity analysis

Augmentation treatment	Cost (£)	QALYs	Incremental cost (£)	Incremental QALYs	ICER (£/QALY)
Probabilistic results^a					
Lithium	4875	1.22809			
AAP	5651	1.23307	776	0.005	155,828
Deterministic results					
Lithium	4876	1.22671			
AAP	5647	1.23399	771	0.01	105,850

a Probabilistic results were generated by averaging the results of 2000 probabilistic runs.

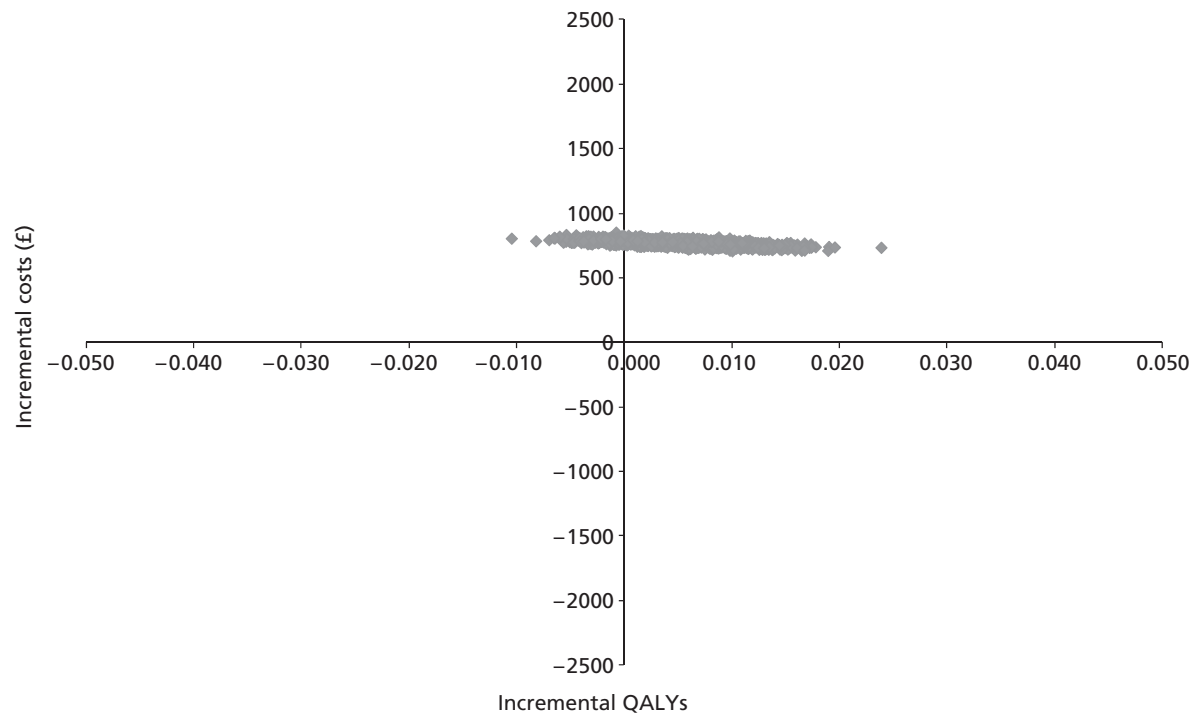


FIGURE 20 Cost-effectiveness plane associated with the probabilistic result of the scenario analysis extending the assumption of class effect based on data from a clinical sensitivity analysis.

In the absence of direct comparison in a RCT, the acute efficacy of treatment was estimated from a MTC carried out as part of the clinical effectiveness review. However, there is a paucity of RCTs considering lithium as an augmentation therapy, and no trial in the required population (failed on two or more antidepressant therapies in the current episode of depression) was identified in the clinical effectiveness review. Consequently, a RCT that matched the prespecified inclusion criteria as closely as possible was used as a surrogate. The RCT ($n=33$) considered lithium augmentation in a population of patients who had failed on one or more previous antidepressant regimes. It is important to note that this trial recruited only 33 patients (17 in the lithium arm and 16 in the SSRI-alone arm) and, as might be expected with such a small sample size, the results are uncertain; this uncertainty is propagated through to the MTC. In addition, it is important to note that the inclusion of some patients with only one previous failure of an antidepressant regime may result in an overall bias in favour of lithium. However, as the relative effect of lithium + SSRI (compared with SSRI alone) was used to inform the MTC, which may be a consistent benefit over antidepressant therapy alone irrespective of number of previous antidepressants, the difference in study populations may have little effect on the results. Nevertheless, the presence or extent of this potential bias remains unknown and it is important to consider the impact of this uncertainty when interpreting the cost-effectiveness results.

The model required data on the outcomes of remission, response and discontinuation. However, data on remission were not reported and the definitions of response varied across the trials included in the MTC. Therefore, a sampling method was used to generate the required probabilities; the treatment effect (change in MADRS score from baseline) with each augmentation therapy was sampled from a distribution of possible effects. The 95% CrIs for the distributions used for each augmentation therapy overlapped (lithium 95% CrI -33.00 to 7.84 ; AAP 95% CrI -30.13 to 7.69) suggesting a non-significant difference in the effect of each treatment. However, the consequence of wider CrIs for lithium led to a higher proportion of patients in the lithium arm achieving remission, a phenomenon that was maintained throughout the probabilistic analysis.

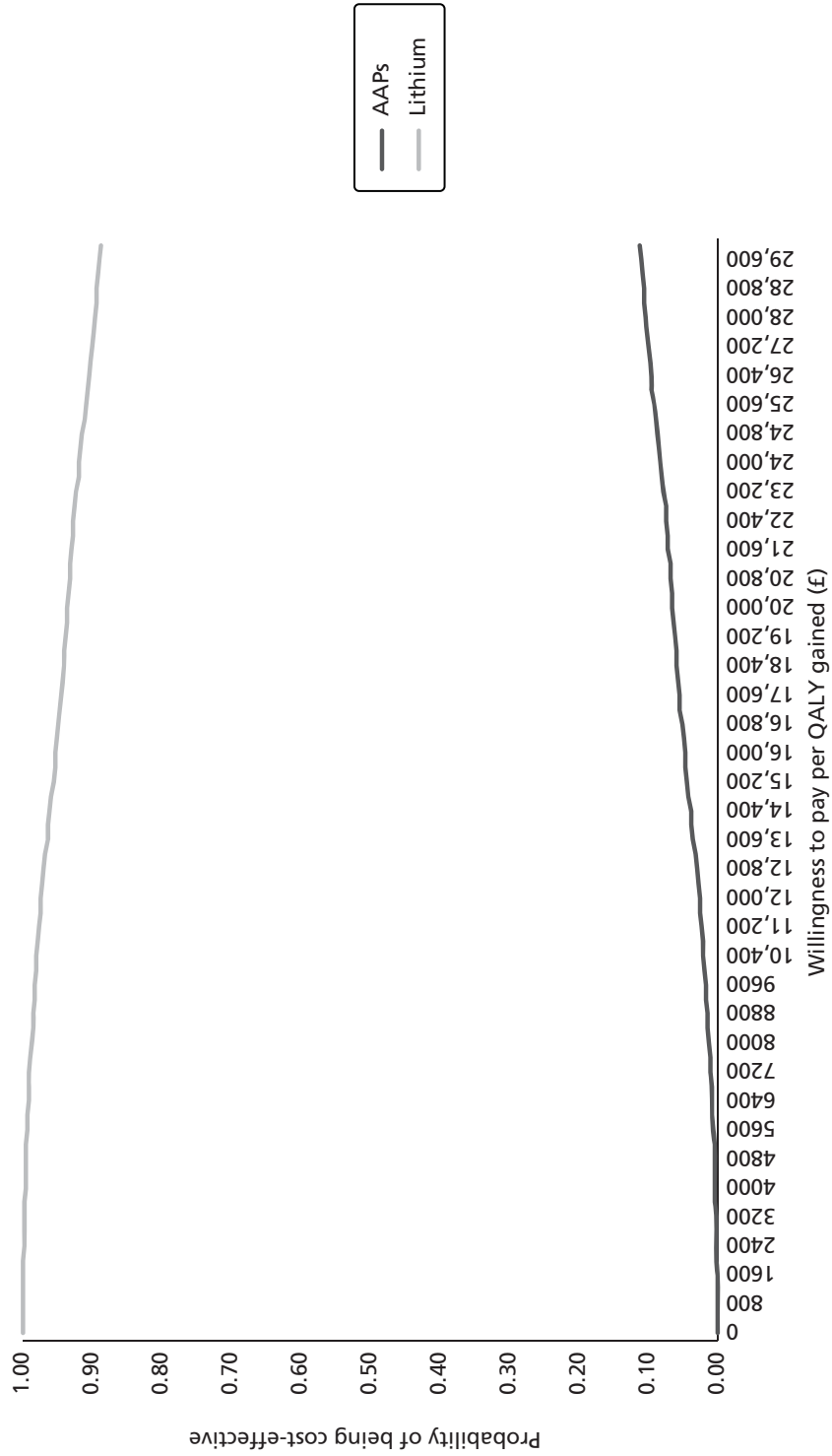


FIGURE 21 Cost-effectiveness acceptability curve associated with the probabilistic result of the scenario analysis extending the assumption of class effect based on data from a clinical sensitivity analysis.

Discontinuation was a key driver of the model, as patients who discontinued their acute therapy were far less likely to experience response or remission. The ORs of discontinuation (vs. SSRI therapy alone) used to inform the economic model were based on the same MTC used to derive the distributions of acute efficacy. Similarly, the 95% Crls for the ORs associated with each augmentation therapy overlapped (lithium 95% Crl 0.13 to 3.32; AAP 95% Crl 0.90 to 1.75) suggesting a non-significant difference in the level of discontinuation. Moreover, the mean OR for discontinuation with lithium (0.92) was almost equal to the lower 95% Crl of the OR for discontinuation with AAP (0.90), resulting in generally more favourable levels of discontinuation with lithium. The OWSA indicated that the acute efficacy (change in MADRS score) and level of discontinuation associated with each treatment are key drivers of the model results. Therefore, it is important to consider the combined effect of the uncertainty surrounding these inputs when interpreting the model's results. Furthermore, in the absence of evidence to inform separate comparisons of each augmentation strategy, the model assumed a class effect for the SSRIs and AAPs. There is evidence of a class effect with the SSRIs from a study by Kroenke *et al.*^{128,129} However, no evidence of a class effect for the AAPs was identified and the results of the clinical sensitivity analysis of class effect were inconclusive. Therefore, this additional uncertainty around the validity of a class effect should be considered when implementing any recommendations based on the economic evaluation presented as part of this review.

Conclusion

The economic evaluation presented as part of this review indicated that, in patients with TRD, augmentation of SSRI therapy with an AAP may be dominated by augmentation with lithium. However, the cost-effectiveness results are predominantly driven by estimates of acute efficacy and discontinuation derived from the assessment of clinical effectiveness. The results of the clinical effectiveness analyses indicate that there is no statistically significant difference between the two augmentation strategies. In addition, the general paucity of RCTs available for augmentation therapy using SSRI as background treatment in patients with TRD results in a high level of uncertainty in both the clinical and economic analyses.

Chapter 6 Overall discussion

This section summarises the principal findings of the clinical effectiveness and cost-effectiveness review of the literature surrounding the augmentation of SSRI therapy with either lithium or an AAP. The key studies identified as part of the clinical effectiveness review and the results of analyses based on these are discussed. Consideration is then given to the interpretation and significance of these results. The results of the cost-effectiveness and QoL literature reviews are also discussed and the development of the de novo model summarised. Any implications of the findings from the de novo economic evaluation are considered in conjunction with any apparent uncertainty.

Statement of principal findings

Description of the identified clinical studies

Initially, 11 RCTs^{30,43,46–53} were identified that considered the augmentation of SSRI therapy in the relevant population (patients who had failed to respond to two or more antidepressant regimens). However, 10 of these RCTs considered the augmentation of SSRI therapy with an AAP.^{43,46–53} The remaining RCT³⁰ considered the augmentation of SSRI or SNRI (venlafaxine) therapy with lithium compared with AAP. None of the identified studies considered augmentation of SSRI therapy with lithium. The primary analysis did not assume a class effect for either the SSRIs or AAPs. Therefore, the single trial identified that considered lithium was ineligible for inclusion in the primary analysis. Consequently, to enable a primary analysis of the relative treatment effect of augmentation with lithium compared with augmentation with an AAP, a surrogate lithium trial was included. The patient population of this trial had failed on one or more antidepressant therapy.

Summary of clinical findings

Primary and secondary analyses of the relative clinical effectiveness were carried out based on the RCTs identified as part of the clinical effectiveness review for the following comparisons:

- SSRI + AAP vs. SSRI alone
- SSRI + lithium vs. SSRI alone
- SSRI + AAP vs. SSRI + lithium.

The primary analyses assumed that the SSRIs and AAPs were associated with different efficacy profiles, whereas the secondary analyses assumed a class effect for both classes of drug.

Selective serotonin reuptake inhibitor plus atypical antipsychotic compared with selective serotonin reuptake inhibitor alone

Of the 12 RCTs identified, 10 were used to inform the comparison of SSRI + AAP with SSRI alone (six in the primary analysis and all 10 in the secondary analysis). The six trials in the primary analysis were all comparisons of fluoxetine (SSRI) plus olanzapine (AAP) with fluoxetine alone. Not all included trials reported data for each outcome assessed. To summarise, six reported response based on the MADRS or HAMD scale, five reported remission based on the MADRS or HAMD scale, four reported MD in MADRS score from baseline and five reported all-cause withdrawal.

For the outcome of response, the use of fixed-effects meta-analysis demonstrated a statistically significant benefit of fluoxetine plus olanzapine compared with fluoxetine alone (OR 1.48; 95% CI 1.13 to 1.94), with moderate heterogeneity ($I^2 = 53\%$; $p = 0.07$). The significance of this benefit was maintained when a random-effects meta-analysis was used to synthesise the data (OR 1.60; 95% CI 1.01 to 2.53). The results of a fixed-effects meta-analysis for the outcome of remission demonstrated a statistically significant increase in remission for patients treated with fluoxetine plus olanzapine compared with fluoxetine alone (OR 1.77; 95% CI 1.27 to 2.47), with no statistical heterogeneity ($I^2 = 0\%$; $p = 0.75$). The outcome of response was the

only outcome considered in the secondary analysis assuming a class effect. The results of both the fixed- and random-effects meta-analyses used to assess this outcome were consistent with the primary analysis; a statistically significant improvement in response was seen with AAP augmentation of baseline therapy.

Meta-analysis of the change in MADRS score from baseline resulted in a statistically significant MD of -2.04 (95% CI -3.25 to -0.82) in favour of treatment with fluoxetine plus olanzapine. However, the level of heterogeneity associated with this analysis was high ($I^2 = 73\%$; $p = 0.01$). It was noted that one of the studies reporting change in MADRS score from baseline (the pooled analysis by Thase *et al.*⁵³) reported a much larger MD compared with other trials included in the meta-analysis. An exploratory meta-analysis excluding data from Thase *et al.*⁵³ resulted in a non-statistically significant MD of -1.15 (95% CI -2.49 to 0.19) in favour of treatment with fluoxetine plus olanzapine, with no statistical heterogeneity ($I^2 = 0\%$; $p = 0.38$).

The results of a fixed-effects meta-analysis of all-cause withdrawal suggested a statistically non-significant reduction in discontinuations with fluoxetine alone compared with fluoxetine plus olanzapine (OR 1.25; 95% CI 0.91 to 1.71), with no statistical heterogeneity ($I^2 = 0\%$; $p = 0.51$).

Selective serotonin reuptake inhibitor plus lithium compared with selective serotonin reuptake inhibitor alone

Only one trial comparing SSRI plus lithium with SSRI alone was identified in the clinical effectiveness review. However, two definitions of response were used, neither of which was consistent with the definition of response used in this review ($\geq 50\%$ reduction in MADRS score from baseline). The results of this trial regarding response indicate a non-significant trend towards improved response with fluoxetine plus lithium compared with fluoxetine alone (a priori definition of response, OR 1.48; 95% CI 0.37 to 5.95; post hoc definition of response, OR 3.85; 95% CI 0.80 to 18.62). The MD in change in MADRS score from baseline between fluoxetine + lithium and fluoxetine alone was -3.79 (95% CI -11.25 to 3.67). This represents a non-significant improvement from baseline score with fluoxetine plus lithium compared with fluoxetine alone. Furthermore, discontinuations were slightly lower with fluoxetine plus lithium (4/17) than with fluoxetine alone (5/16).

Selective serotonin reuptake inhibitor plus atypical antipsychotic compared with selective serotonin reuptake inhibitor plus lithium

Of the 12 trials identified in the clinical effectiveness review, 10 were used to inform the comparison of SSRI + AAP compared with SSRI + lithium (seven in the primary analysis and all 10 in the secondary analysis). Of the seven trials included in the primary analysis, six were comparisons of fluoxetine plus olanzapine with fluoxetine alone and one looked at fluoxetine plus lithium compared with lithium alone. Not all included trials reported data for each outcome assessed. To summarise, seven trials reported data for response based on MADRS or HAMD score, five trials reported data for MD in MADRS score from baseline and six trials reported data on all-cause withdrawal.

For the outcome of response (using data from the post hoc definition of response from the lithium trial), a statistically non-significant trend in favour of augmentation with lithium was detected (OR 4.15; 95% CrI 0.25 to 20.34). However, it is important to note that the definition of response used in the trial comparing lithium plus fluoxetine with fluoxetine alone differed from that used in the trials comparing fluoxetine plus olanzapine with fluoxetine alone. The MD in the change in MADRS score from baseline determined from the MTC also suggested a non-significant trend in favour of augmentation with lithium (OR -1.47 ; 95% CrI -9.10 to 6.41). Furthermore, the results of the MTC with respect to all-cause withdrawal indicated a non-significant benefit for patients augmented with lithium (OR 0.74; 95% CrI 0.10 to 2.66).

Discussion of clinical findings

The available clinical effectiveness data informing the comparison of SSRI + AAP with SSRI + lithium in the primary analysis was based on fluoxetine + olanzapine^{43,49,51-53} compared with fluoxetine + lithium.⁵⁹ The results from the MTC of the star-shaped network demonstrate no significant differences between treatment regimens for any of the outcomes assessed. A non-significant trend in benefit was observed for the

lithium-based augmentation strategy compared with fluoxetine + olanzapine for response, mean change in MADRS from baseline, and fewer discontinuations. The results of the MTC also demonstrated a non-significant trend in favour of the olanzapine-based augmentation strategy compared with fluoxetine + lithium for fewer adverse events. However, care should be taken when interpreting non-significant results.

When the results of the MTC are compared with the individual results for the pairwise meta-analyses, there is general agreement with the results obtained when SSRI is used as the baseline. This suggests a reasonable fit of the modelling approach used within the MTC. The radiating star shape of the network means that only the trials providing the results from the pair-wise meta-analyses are providing the results within the MTC. The results for the lithium-based augmentation strategy compared with SSRI alone in the MTC tend to have wider 95% CrIs than the 95% CIs provided from the single trial informing this comparison.⁵⁹ This is likely to be due to the random-effects model tending to be the preferred model for the MTC outcomes assessed.

The results for both the pair-wise meta-analysis and MTC estimates of SSRI + AAP compared with SSRI alone showed a statistically significant benefit in favour of augmentation with AAP for the outcomes of response and mean change in MADRS score. The equivalent results for SSRI + lithium compared with SSRI alone showed a statistically non-significant trend in favour of augmentation with lithium. The results for lithium augmentation could be considered inconclusive, although it should be noted that they are based on data from only a small subgroup of patients in one RCT⁵⁹ and other publications have reported results demonstrating lithium to be an effective augmentation strategy.^{60,65,66} A recent meta-analysis by Crossley *et al.*⁶⁰ included 10 RCTs and demonstrated a statistically significant benefit in terms of response rate with lithium augmentation compared with placebo (OR 3.11; 95% CI 1.80 to 5.37). This meta-analysis is not entirely comparable with the patient population under review in the current research because it included patients with bipolar disorder, patients on various antidepressants and patients with a minimum of one previous antidepressant failure. However, the Crossley *et al.*⁶⁰ meta-analysis does provide strong evidence to suggest that lithium is an effective augmentation agent in TRD and thus is supportive of the results from the MTC.

Relapse rates and mortality were prespecified outcomes of interest for this review; however, none of the trials included reported comparable mortality or relapse rate data. In addition, extremely limited subgroup data, if any, were reported for the trials included in the clinical effectiveness review, and no trial reported suitable subgroup data for the prespecified subgroup analyses.

The trials included in this review were validated against the trials included in the NICE clinical guideline for depression in adults (CG90)¹⁶ and those included in two separate systematic reviews; one for the comparison of AAP augmentation with placebo in MDD¹¹ and the other for lithium augmentation compared with placebo in TRD.⁶⁰ All three publications included additional trials compared with this review, although none of the additional trials were found to be suitable for inclusion. Moreover, each trial was excluded for multiple reasons based on the inclusion criteria for this review. Mostly, the additional RCTs did not meet the following criteria: the augmentation of SSRI as baseline therapy,⁶⁷⁻⁸¹ a 4-week minimum duration of treatment;⁷⁴⁻⁸⁰ or two or more failures of antidepressant therapy in the current episode of depression⁶⁷⁻⁷³ (it is unclear how many trials in Crossley *et al.*⁶⁰ are affected by this criterion). The results of both the Crossley *et al.*⁶⁰ meta-analysis evaluating lithium compared with placebo augmentation and the Nelson *et al.*¹¹ review of AAP compared with placebo augmentation demonstrated that the respective augmentation agents were statistically significantly more effective than placebo at achieving treatment response in people with MDD or TRD.

The sensitivity analysis based on a class effect (sensitivity analysis 1) allowed the inclusion of two additional AAPs (aripiprazole and quetiapine) in the MTC. For the outcomes assessed (response) the non-significant trend favouring lithium augmentation observed in the primary analysis is diminished (mean OR changes from 4.15 to 1.07) and remains non-significant. This could be supportive evidence for no clinically meaningful

difference between the two augmentation strategies, a reflection of the lack of information available for the comparison, or an indication that the assumption of a class effect is flawed and that there is a difference between the individual treatments within a class. With regards to the last concern, the trial providing information on quetiapine in the MTC³⁰ demonstrated a non-significant trend in favour of a quetiapine-based augmentation strategy compared with lithium-based augmentation (OR 1.25; 95% CI 0.74 to 2.12). The result of this trial is thus in contrast with the result of the MTC and therefore suggests a difference in treatment effect between the different AAPs.

Summary of the cost-effectiveness review

No economic evaluations were identified that considered the relative cost-effectiveness of augmentation of SSRI therapy with lithium or with an AAP. However, four economic evaluations were identified that included a population of people with depression who were treatment resistant. Each study highlighted issues surrounding the economic evaluation of TRD, particularly around the long-term consequences of treatment and the paucity of data available in this area. A variety of approaches were taken to assess the cost-effectiveness of various interventions (mostly SSRI therapy). The hybrid approach used by Simpson *et al.*⁸⁹ was considered to be most relevant to the current decision problem and was used to inform the structure of the de novo economic model.

Summary of quality-of-life review

A systematic review of the QoL literature was carried out to identify:

- HSUV to inform the economic model
- any issues that positively or negatively impact on the QoL of depressed patients.

Five studies reporting utility values^{91,102–105} by severity of depression or level of response were identified. Of these, only one study by Sapin *et al.*¹⁰⁵ met the criteria for elicitation and valuation outlined in the NICE methods guide.¹⁰⁸ The study was carried out in France but used UK general public valuation scores. Sapin *et al.*¹⁰⁵ reported utility by level of treatment response (remission, response and non-response) and by severity of depression (mild, moderate and severe). Therefore, it was these utility values that were used to inform the QALY calculations in the de novo economic model.

A further 12 studies were identified that considered the overall QoL in depressed patients.^{30,43,46–53,59} There is evidence from these studies to suggest that treatments for depression are effective in improving the QoL of patients, particularly if treatment is sustained. In addition, the QoL evidence base indicates that pharmacological treatment of depression does not result in any improvement of physical outcomes, and physical therapies do not result in any improvement of psychological symptoms. This, in turn, suggests that the physical and psychological domains of QoL are independent of each other in determining the overall QoL of a depressed patient.

Summary of economic evaluation

A hybrid economic model was constructed to simulate the clinical and economic consequences of augmenting an SSRI with either lithium or an AAP in the treatment of TRD. The model considered outcomes from the perspective of the NHS over a 1-year time horizon. The model consists of two distinct components: a decision tree (8-week time horizon) and a Markov component (10-month time horizon). A hybrid model was chosen, as this facilitated capturing the granularity of the acute treatment phase and also accounted for the patient progression within 1 year.

The model was constructed around the level of response to acute treatment (assessed by changes in MADRS score) and the extent to which this response (or otherwise) is maintained. The acute efficacy and discontinuation associated with each treatment was obtained from a MTC carried out as part of the clinical effectiveness review. Discontinuation was a key driver of the model, as patients who discontinued their acute therapy were far less likely to experience response or remission.

Furthermore, as a result of the paucity of long-term data, progression within the Markov component of the model was assumed to be independent of treatment regimen. The progression of patients who did not respond to acute treatment (and patients who relapsed during the Markov component of the model) was simplified such that these patients were assumed not to benefit from further augmentation treatment and received a standard package of care. It is acknowledged that, in reality, patients who do not respond to (or relapse following) acute augmentation therapy are likely to experience some benefit from further treatment (e.g. change in augmentation treatment, ECT, etc.). However, these follow-on patient pathways are diverse and there is a paucity of data to inform them, and, therefore, the decision was made to simplify this aspect of patient care. In addition, while non-responding patients were exposed to the probability of 'spontaneous' remission or response, in accordance with expert clinical opinion, patients who relapsed were assumed to enter an absorbing health state.

Summary of economic findings

The base-case results indicated that augmentation of SSRI therapy with an AAP is dominated by augmentation of an SSRI with lithium in patients with TRD. Furthermore, PSA suggested that the overall health benefit obtained with AAP augmentation was consistently less than the overall health benefit obtained with lithium augmentation; none of the probabilistic runs resulted in an incremental QALY gain for AAP augmentation over lithium augmentation.

However, the cost-effectiveness results are predominantly driven by estimates of acute efficacy and discontinuation derived from the assessment of clinical effectiveness. The results of the clinical effectiveness analyses indicate that there is no statistically significant difference between the two augmentation strategies. In addition, the general paucity of RCTs considering the augmentation of SSRI therapy in a TRD population results in a high level of uncertainty in the clinical effectiveness results. In particular, there are a limited number of robust data available to inform the clinical effectiveness of lithium augmentation. This uncertainty is carried through to the economic analysis and is potentially perpetuated by the structural assumptions made. Although every effort has been made to ensure that the model structure is in line with current clinical expectations, it is acknowledged that the natural history of TRD remains somewhat unpredictable.

Strengths and limitations and uncertainty of the assessment

The main strengths of this review were the:

- systematic identification of studies used to inform the clinical effectiveness and cost-effectiveness analyses
- robust methodology used to synthesise the clinical efficacy evidence in meta-analysis and MTC
- consultation with clinical experts throughout the development process of the de novo economic evaluation
- use of sampling to synthesise the clinical effectiveness evidence within the de novo economic model.

To systematically identify clinical effectiveness and cost-effectiveness evidence, all relevant databases were searched from inception. In addition, the CCDAN was contacted for access to their study registries. The website ClinicalTrials.gov was searched to identify relevant ongoing clinical trials and clinical experts in the relevant therapy areas were contacted for details of published or unpublished trials.

This evaluation of the available literature constitutes the most comprehensive research into augmentation therapy for TRD. Although few RCTs were identified, the identified RCTs represent the best available data on which to base a comparison of augmentation of SSRI with lithium or AAP. In the absence of a direct comparison in a RCT, use of a MTC has been shown to be one of the most reliable methods for indirectly

comparing treatments estimates.³⁹ By using a network of connected trials, in which a new trial has to include a comparator already existing in the network, valid estimates of relative efficacy can be obtained.³³⁻³⁵

The key criterion for a valid network is to be populated with a set of RCTs that are as similar as possible. A strength of this review is the rigour with which the inclusion/exclusion criteria were applied to obtain as homogeneous as possible set of trials. It is unfortunate that, for the lithium comparator, the nearest trial identified to the type of trial required was a less than perfect match for one inclusion criterion.⁵⁹ However, the researchers (VH, SB, SJE) agreed that it was the best approximation of the required trial available. In addition, it was unfortunate the RCTs identified for the primary analysis only provided sufficient data for a 'stepwise' indirect comparison of the data, i.e. A compared with B compared with C, without the possibility of drawing further strength and cohesion in the network that a trial of A compared with C would have provided.

To ensure the model reflected the management of unipolar depression in the UK, two clinical advisors (both of whom are practising psychiatrists with experience of managing TRD in either primary or secondary care) were consulted to validate model assumptions throughout the model development process. Information from experts was obtained through teleconferences and standardised questionnaires.

The use of sampling to estimate the acute efficacy of each augmentation strategy was necessary because of the paucity of usable information from the identified clinical effectiveness studies. However, the use of this technique circumvented the need for a consistent definition of response or remission across studies.

The most significant limitation to this review was the absence of a direct RCT comparing augmentation of SSRI therapy with lithium to augmentation with an AAP. In addition, the general paucity of RCTs considering each individual comparator in the required population was a significant limitation, particularly with respect to lithium augmentation for which no RCTs were available in the required population, which necessitated the inclusion of a surrogate trial.⁵⁹

The single RCT identified comparing lithium augmentation of SSRI to SSRI alone was the main source of uncertainty in the primary clinical analysis. This is because the sample size of this trial was small, with only 33 patients randomised. In addition, this trial was the only link for the stepwise indirect comparison of AAP augmentation and lithium augmentation. Furthermore, the observed withdrawal profile obtained from this small RCT was counter to prior clinical expectation, in that fewer withdrawals were seen with SSRI + lithium than with SSRI alone. However, this may be a reflection of the small size of the trial; in a small trial, one or two patients having a different outcome in a study arm can have a dramatic effect on the OR between two study groups.

Restriction of the scope of background therapy for which augmentation treatment was considered was a further limitation of this review: consideration of other background therapies may have provided additional linking studies to enable the comparison of the treatments of interest.

Regarding the economic component to this review, the use of non-standard methodology that has not been previously validated was a limitation. This limitation was further enhanced by the uncertainty in the clinical data, which resulted in counterintuitive cost-effectiveness results; the treatment associated with the most uncertainty displayed the most benefit under the sampling methodology. This is because the threshold of benefit (in terms of MADRS score) is fixed. Therefore, treatments associated with more uncertainty (which are more likely to achieve a wider range of scores) are more likely to achieve higher rates of remission and response as a result of random chance. However, treatments associated with more uncertainty are also more likely to have more severe levels of non-response. This uncertainty was not captured in the methodology used, which samples treatment effect and categorises the samples around a fixed threshold of benefit.

Finally, the paucity of data to inform long-term patient progression limited the economic modelling of medium to long-term consequences of each individual therapy.

The overall results of this review are highly uncertain with regards to which augmentation strategy is most clinically effective and so which is the most cost-effective. However, the review does highlight the requirement for a direct head-to-head clinical trial to assess SSRI + lithium and SSRI + AAP in patients with TRD to enable more definitive conclusions to be drawn.

Chapter 7 Conclusions

The results of this review support the conclusion that augmentation of SSRIs with lithium or an AAP is likely to be beneficial in people with TRD, defined as failure to respond to two or more antidepressants in the current episode of depression. However, based on the limited number of RCTs identified in this research, the clinical evaluation suggests there is no statistically significant difference between the two augmentation strategies. There is a general paucity of trial data available in patients with TRD for SSRI + lithium and SSRI + AAP.

The cost-effectiveness results suggest that augmentation with lithium is cheaper and more effective than augmenting with AAP. However, the results are not definitive because the model is sensitive to the clinical effectiveness parameters of discontinuation and treatment response. The cost-effectiveness of SSRI + lithium and SSRI + AAP will need to be reconsidered if data from a direct comparison in a RCT or additional trial data that would inform the MTC become available.

Implications for service provision

The results of this research suggest that augmentation of SSRIs with lithium or an AAP is likely to be beneficial in people with TRD, defined as failure to respond to two or more antidepressants in the current episode of depression. However, the overall results of this research are highly uncertain with regards to which augmentation strategy is most clinically effective and so which is likely to be cost-effective. This report highlights the requirement for further research to assess SSRI + lithium and SSRI + AAP in patients with TRD to enable a more definitive conclusion to be drawn.

Suggested research priorities

A RCT is needed, comparing SSRI + lithium and SSRI + AAP in patients with TRD. The research should collect all relevant outcomes, which are response, remission, and discontinuation rates (especially treatment related) in the short term, as well as collecting long-term outcome (e.g. relapse rate) data during maintenance treatment. Adverse events and QoL data should also be prioritised for collection as part of the research. In addition, as part of this research, it has become clear that other antidepressants are also used as baseline therapy for augmentation in TRD, such as tricyclic antidepressants and SNRIs. The current research could be expanded to include these treatments.

Acknowledgements

Thanks to the following people for their assistance in the production of this report:

- Samantha Barton, Health Technology Assessment Analyst, BMJ Technology Assessment Group, London, UK.
- Lili Ng, Developer, BMJ Technology, London, UK.
- Professor Philip J Cowen, MRC Clinical Scientist and Professor of Psychopharmacology, Warneford Hospital, Oxford, UK.
- Dr Luiz Dratcu, Consultant Psychiatrist, South London & Maudsley NHS Foundation Trust, London, UK.
- Dr Hamish McAllister-Williams, Reader in Clinical Psychopharmacology & Honorary Consultant Psychiatrist, Newcastle University, Newcastle upon Tyne, UK.
- Dr Andrea L Malizia, Consultant Senior Lecturer (Clinical Psychopharmacology and Neurostimulation), University of Bristol, Bristol, UK.
- Ifigenia Mavranouzouli, Senior Health Economist, National Collaborating Centre for Mental Health, University College London, London, UK.
- The Cochrane Collaboration Depression, Anxiety and Neurosis Review Group (CCDAN), University of Bristol, Bristol, UK.

Thanks also to those manufacturers and authors of key trials who responded to requests for additional data.

Contributions of authors

Steve Edwards Project lead: supervised the production of the final report; data validation and validation of analyses; critical appraisal of the clinical evidence; and critical appraisal of the economic evidence.

Victoria Hamilton Critical appraisal of the clinical evidence and drafted the clinical sections.

Leo Nherera Critical appraisal of the economic evidence and drafted the economic sections and scientific summary.

Nicola Trevor Critical appraisal of the economic evidence and drafted part of the economic sections, the discussion and overall conclusions.

All authors read and commented on draft versions of the report.

References

1. Jenkins E, Goldner EM. Approaches to understanding and addressing treatment-resistant depression: a scoping review. *Depres Res Treat* 2012. URL: www.hindawi.com/journals/drt/2012/469680/ (accessed 23 April 2013).
2. Tierney JG. Treatment-resistant depression: managed care considerations. *J Manag Care Pharm* 2007;**13**(Suppl.):2–7.
3. National Institute for Health and Care Excellence (NICE). *Depression: the treatment and management of depression in adults*. CG23. London: NICE; 2004.
4. Berlim MT, Turecki G. What is the meaning of treatment resistant/refractory depression (TRD)? A systematic review of randomized trials. *Eur Neuropsychopharmacol* 2007;**17**:696–707. <http://dx.doi.org/10.1016/j.euroneuro.2007.03.009>
5. Bauer M, Whybrow PC, Angst J, Versiani M, Möller HJ. World Federation of Societies of Biological Psychiatry (WFSBP) guidelines for biological treatment of unipolar depressive disorders, part 1: acute and continuation treatment of major depressive disorder. *World J Biol Psychiatry* 2002;**3**:5–43. <http://dx.doi.org/10.3109/15622970209150599>
6. Kupfer DJ. Long-term treatment of depression. *J Clin Psychiatry* 1991;**52**(Suppl. 5):28–34. [http://dx.doi.org/10.1016/S0165-0327\(00\)00357-8](http://dx.doi.org/10.1016/S0165-0327(00)00357-8)
7. Musselman DL, Evans DL, Nemeroff CB. The relationship of depression to cardiovascular disease. *Arch Gen Psychiatry* 1998;**55**:580–92. <http://dx.doi.org/10.1001/archpsyc.55.7.580>
8. Bostwick JM, Pankratz VS. Affective disorder and suicide risk: a re-examination. *Am J Psychiatry* 2000;**157**:1925–32. <http://dx.doi.org/10.1176/appi.ajp.157.12.1925>
9. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders (DSM-IV-TR)*. 4th edn, text revision. Arlington State, VA: American Psychiatric Association; 1994. URL: <http://dsm.psychiatryonline.org/book.aspx?bookid=22> (accessed 31 May 2012).
10. World Health Organization (WHO). *International classification of diseases*. URL: www.who.int/classifications/icd/en/ (accessed 31 May 2012).
11. Nelson JC, Papakostas GI. Atypical antipsychotic augmentation in major depressive disorder: a meta-analysis of placebo-controlled randomized trials. *Am J Psychiatry* 2009;**166**:980–91. <http://dx.doi.org/10.1176/appi.ajp.2009.09030312>
12. Rush AJ, Kraemer HC, Sackeim HA, Fava M, Trivedi MH, Frank E, et al. Report by the ACNP Task Force on response and remission in major depressive disorder. *Neuropsychopharmacology* 2006;**31**:1841–53. <http://dx.doi.org/10.1038/sj.npp.1301131>
13. Hamilton M. A rating scale for depression. *J Neurol Neurosurg Psychiatry* 1960;**23**:56–62. <http://dx.doi.org/10.1136/jnnp.23.1.56>
14. Montgomery SA, Åsberg M. A new depression scale designed to be sensitive to change. *Br J Psychiatry* 1979;**134**:382–9. <http://dx.doi.org/10.1192/bjp.134.4.382>
15. Zimmerman M, Posternak MA, Chelminski I. Defining remission on the Montgomery–Åsberg Depression Rating Scale. *J Clin Psychiatry* 2004;**65**:163–8. <http://dx.doi.org/10.4088/JCP.v65n0204>
16. National Institute for Health and Care Excellence (NICE). *Depression: the treatment and management of depression in adults (update)*. CG90. London: NICE; 2009.
17. Singleton N, Lee A, Meltzer H. *Psychiatric morbidity among adults living in private households*. Technical Report. London: Office for National Statistics; 2007.

18. Thomas CM, Morris S. Cost of depression among adults in England in 2000. *Br J Psychiatry* 2003;**183**:514–19. <http://dx.doi.org/10.1192/bjp.183.6.514>
19. British Medical Association and Royal Pharmaceutical Society of Great Britain. *British national formulary*. No. 63, March 2012. London: BMA and RPS; 2012.
20. Anderson IM, Ferrier IN, Baldwin RC, Cowen PJ, Howard L, Lewis G, *et al*. Evidence-based guidelines for treating depressive disorders with antidepressants: a revision of the 2000 British Association for Psychopharmacology guidelines. *J Psychopharmacol* 2008;**22**:343–96. <http://dx.doi.org/10.1177/0269881107088441>
21. Greenberg P, Corey-Lisle PK, Birnbaum H, Marynchenko M, Claxton A. Economic implications of treatment-resistant depression among employees. *Pharmacoeconomics* 2004;**22**:363–73. <http://dx.doi.org/10.2165/00019053-200422060-00003>
22. McCrone P, Dhanasiri S, Patel A, Knapp M, Lawton-Smith S. *Paying the price: the cost of mental health care in England to 2026*. London: The King's Fund; 2008.
23. Mental Health Foundation. *New Government figures reveal the economic burden of depression has risen to £9bn a year*. URL: www.mentalhealth.org.uk/our-news/news-archive/2010/2010-11-22/?view=Standard (accessed May 2012).
24. The NIHR Evaluation, Trials and Studies Coordinating Centre (NETSCC), HTA. *Management of treatment-resistant depression*. www.hta.ac.uk/protocols/201000300001.pdf (accessed May 2012).
25. Edwards S, Hamilton V, Nherera L, Barton S, Trevor N. Lithium or an atypical anti-psychotic in the management of treatment resistant depression: systematic review and economic evaluation. *PROSPERO* 2011:CRD42011001464. URL: www.crd.york.ac.uk/PROSPERO/display_record.asp?ID=CRD42011001464 (accessed 31 May 2012).
26. Schulz KF, Altman DG, Moher D, for the CONSORT Group. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. *Int J Surg* 2011;**9**:672–7. <http://dx.doi.org/10.1016/j.ijsu.2011.09.004>
27. Higgins JPT, Green S, editors. *Cochrane handbook for systematic reviews of interventions*. Version 5.1.0 (updated March 2011). The Cochrane Collaboration; 2011. URL: www.cochrane-handbook.org (accessed 31 May 2012).
28. Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. *J Natl Cancer Inst* 1959;**22**:719–48.
29. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials* 1986;**7**:177–88. [http://dx.doi.org/10.1016/0197-2456\(86\)90046-2](http://dx.doi.org/10.1016/0197-2456(86)90046-2)
30. Franco MM, Figueira ML, Petrov P, Rihmer Z, Vavrusov L, Dencker-Vansvik E, *et al*. Quetiapine XR monotherapy and quetiapine XR + ongoing antidepressant vs lithium + ongoing AD for Stage II treatment-resistant major depressive disorder. *Eur Neuropsychopharmacol* 2010;**20**(Suppl. 3):347.
31. Bauer M, Dell'Osso L, Kasper S, Pitchot W, Dencker-Vansvik E, Jörgensen L, *et al*. Quetiapine XR monotherapy, quetiapine XR + ongoing antidepressants and lithium + ongoing antidepressants in patients with treatment-resistant major depressive disorder. *Int J Neuropsychopharmacol* 2010;**13**(Suppl. 1):143.
32. Montgomery S, Bauer M, Dell'Osso L, Kasper S, Pitchot W, Dencker-Vansvik E, *et al*. Quetiapine XL monotherapy, quetiapine XL + ongoing antidepressants and lithium + ongoing antidepressants in patients with treatment-resistant major depressive disorder (MDD): a UK perspective. Poster presented at the British Association for Psychopharmacology, Harrogate, North Yorkshire, UK, 25–28 July 2010.

33. Ades AE. A chain of evidence with mixed comparisons: models for multiparameter synthesis and consistency of evidence. *Stat Med* 2003;**22**:2995–3016.
34. Lu G, Ades AE. Combination of direct and indirect evidence in mixed treatment comparisons. *Stat Med* 2004;**23**:3105–24. <http://dx.doi.org/10.1002/sim.1875>
35. Caldwell DM, Ades AE, Higgins JPT. Simultaneous comparison of multiple treatments: combining direct and indirect evidence. *BMJ* 2005;**331**:897–900. <http://dx.doi.org/10.1136/bmj.331.7521.897>
36. Spiegelhalter DJ, Best NG, Carlin BP, Van Der Linde A. Bayesian measures of model complexity and fit. *J Roy Statist Soc B* 2002;**64**:583–639. <http://dx.doi.org/10.1111/1467-9868.00353>
37. Dempster AP. The direct use of likelihood for significance testing. *Stat Comp* 1997;**7**:247–52.
38. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ* 2003;**327**:557–60. <http://dx.doi.org/10.1136/bmj.327.7414.557>
39. Edwards SJ, Clarke MJ, Wordsworth S, Borrill J. Indirect comparisons of treatments based on systematic reviews of randomised controlled trials. *Int J Clin Pract* 2009;**63**:841–54. <http://dx.doi.org/10.1111/j.1742-1241.2009.02072.x>
40. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ* 1997;**315**:629–34. <http://dx.doi.org/10.1136/bmj.315.7109.629>
41. Sterne JAC, Egger M. Funnel plots for detecting bias in meta-analysis: guidelines on choice of axis. *J Clin Epidemiol* 2001;**54**:1046–55.
42. Bin W, Mei H, Po L. Fluoxetine and olanzapine in the treatment of refractory depression. *Shandong Arch Psychiatry* 2006;**19**:87–9.
43. Feng H, Meifang C, Yunhai T. Efficacy of olanzapine and fluoxetine combination therapy in treatment-resistant depression. *Zhejiang Pract Med* 2008;**13**:117–18.
44. Li H, Zhang Y, Zhang Y. Double-blind study of fluoxetine augmented with olanzapine in the treatment of treatment-resistant depression. *Shandong Arch Psychiatry* 2006;**19**:85–6.
45. Wang XH, Guo X. Effectiveness and safety of olanzapine combined with fluoxetine for refractory depression: a systematic review. *Chin J Evid Med* 2010;**10**:1102–9.
46. Mattingly G, Ilivicky H, Canale J, Anderson R. Quetiapine combination for treatment-resistant depression. *Annual Meeting of the American Psychiatric Association, New Research Abstracts (Toronto, Canada, 20–25 May, 2006)*. Washington, DC: APA; 2006.
47. Berman RM, Marcus RN, Swanink R, McQuade RD, Carson WH, Corey-Lisle PK, *et al*. The efficacy and safety of aripiprazole as adjunctive therapy in major depressive disorder: a multicenter, randomized, double-blind, placebo-controlled study. *J Clin Psychiatry* 2007;**68**:843–53. <http://dx.doi.org/10.4088/JCP.v68n0604>
48. Berman RM, Fava M, Thase ME, Trivedi MH, Swanink R, McQuade RD, *et al*. Aripiprazole augmentation in major depressive disorder: a double-blind, placebo-controlled study in patients with inadequate response to antidepressants. *CNS Spectr* 2009;**14**:197–206.
49. Corya SA, Williamson D, Sanger TM, Briggs SD, Case M, Tollefson G, *et al*. A randomized, double-blind comparison of olanzapine/fluoxetine combination, olanzapine, fluoxetine, and venlafaxine in treatment-resistant depression. *Depress Anxiety* 2006;**23**:364–72. <http://dx.doi.org/10.1002/da.20130>
50. Marcus RN, McQuade RD, Carson WH, Hennicken D, Fava M, Simon JS, *et al*. The efficacy and safety of aripiprazole as adjunctive therapy in major depressive disorder: a second multicenter, randomized, double-blind, placebo-controlled study. *J Clin Psychopharmacol* 2008;**28**:156–65. <http://dx.doi.org/10.1097/JCP.0b013e31816774f9>

51. Shelton RC, Tollefson GD, Tohen M, Stahl S, Gannon KS, Jacobs TG, *et al.* A novel augmentation strategy for treating resistant major depression. *Am J Psychiatry* 2001;**158**:131–4. <http://dx.doi.org/10.1176/appi.ajp.158.1.131>
52. Shelton RC, Williamson DJ, Corya SA, Sanger TM, Van Campen LE, Case M, *et al.* Olanzapine/fluoxetine combination for treatment-resistant depression: a controlled study of SSRI and nortriptyline resistance. *J Clin Psychiatry* 2005;**66**:1289–97. <http://dx.doi.org/10.4088/JCP.v66n1012>
53. Thase ME, Corya SA, Osuntokun O, Case M, Henley DB, Snager TM, *et al.* A randomized, double-blind comparison of olanzapine/fluoxetine combination, olanzapine, and fluoxetine in treatment-resistant major depressive disorder. *J Clin Psychiatry* 2007;**68**:224–36. <http://dx.doi.org/10.4088/JCP.v68n0207>
54. Shelton RC, Tohen M, Stahl S, Jacobs T, Gannon KS, Tollefson GD. The study of olanzapine plus fluoxetine in treatment-resistant major depressive disorder without psychotic features. *Schizophr Res* 2000;**1**:Special Issue. [http://dx.doi.org/10.1016/S0920-9964\(00\)90772-2](http://dx.doi.org/10.1016/S0920-9964(00)90772-2)
55. Tohen M, Shelton R, Tollefson GD, Stahl S, Jacobs T, Gannon KS, *et al.* Olanzapine plus fluoxetine: double-blind and open-label results in treatment-resistant major depressive disorder. *Eur Neuropsychopharmacol* 1999;**9**(Suppl. 5):S246. [http://dx.doi.org/10.1016/S0924-977X\(99\)80224-7](http://dx.doi.org/10.1016/S0924-977X(99)80224-7)
56. Tollefson GD, Shelton R, Tohen M, Stahl S, Jacobs T, Buras W, *et al.* Efficacy of olanzapine, fluoxetine and combination therapy in treatment-resistant major depressive disorder without psychotic features. 11th European College of Neuropsychopharmacology Congress, Paris, France, 31 October to 4 November 1998.
57. Dunner DL, Amsterdam JD, Shelton RC, Loebel A, Romano SJ. Efficacy and tolerability of adjunctive ziprasidone in treatment-resistant depression: a randomized, open-label, pilot study. *J Clin Psychiatry* 2007;**68**:1071–7. <http://dx.doi.org/10.4088/JCP.v68n0714>
58. ClinicalTrials.gov. *Ziprasidone augmentation of SSRIs for patients with major depressive disorder (MDD) that do not sufficiently respond to treatment with SSRIs*. URL: <http://clinicaltrials.gov/ct2/show/NCT00633399> (accessed May 2012).
59. Katona CL, Abou-Saleh MT, Harrison DA, Nairac BA, Edwards DR, Lock T, *et al.* Placebo-controlled trial of lithium augmentation of fluoxetine and lofepramine. *Br J Psychiatry* 1995;**166**:80–6. <http://dx.doi.org/10.1192/bjp.166.1.80>
60. Crossley NA, Bauer M. Acceleration and augmentation of antidepressants with lithium for depressive disorders: two meta-analyses of randomized, placebo-controlled trials. *J Clin Psychiatry* 2007;**68**:935–40. <http://dx.doi.org/10.4088/JCP.v68n0617>
61. Soares HP, Daniels S, Kumar A, Clarke M, Scott C, Swann S, *et al.* Bad reporting does not mean bad methods for randomised trials: observational study of randomised controlled trials performed by the Radiation Therapy Oncology Group. *BMJ* 2004;**328**:22–5.
62. Thase ME, Trivedi MH, Nelson JC, Fava M, Swanick R, Tran QV, *et al.* Examining the efficacy of adjunctive aripiprazole in major depressive disorder: a pooled analysis of 2 studies. *J Clin Psychiatry* 2008;**10**:440–7.
63. Abrams K. Interpreting tau: OR Scale. Presented at the Indirect and Mixed Treatment Comparisons Course, University of Leicester, UK, 19–21 September 2007.
64. Zimmerman M, Posternak A, Chelminski I. Derivation of a definition of remission on the Montgomery–Åsberg depression rating scale corresponding to the definition of remission on the Hamilton rating scale for depression. *J Psychiatr Res* 2004;**38**:577–82. <http://dx.doi.org/10.1016/j.jpsychires.2004.03.007>

65. Bauer M, Dopfmer S. Lithium augmentation in treatment-resistant depression: meta-analysis of placebo-controlled studies. *J Clin Psychopharmacol* 1999;**19**:427–34. <http://dx.doi.org/10.1097/00004714-199910000-00006>
66. Bauer M, Adli M, Bschor T, Pilhatsch M, Pfennig A, Sasse J, *et al.* Lithium's emerging role in the treatment of refractory major depressive episodes: augmentation of antidepressants. *Neuropsychobiology* 2010;**62**:36–42. <http://dx.doi.org/10.1159/000314308>
67. Mahmoud RA, Pandina G, Turkoz I, Kosik-Gonzalez C, Canuso CM, Kujawa MJ, *et al.* Risperidone for treatment-refractory major depressive disorder. *Ann Intern Med* 2007;**147**:593–602. <http://dx.doi.org/10.7326/0003-4819-147-9-200711060-00003>
68. Reeves H, Batra S, May RS, Zhang R, Dahl DC, Li X. Efficacy of risperidone augmentation to antidepressant in the management of suicidality in major depressive disorder: a randomized, double-blind, placebo controlled pilot study. *J Clin Psychiatry* 2008;**69**:1228–336.
69. Keitner GI, Garlow SJ, Ryan CE, Ninan PT, Solomon DA, Nemeroff CB, *et al.* A randomized, placebo-controlled trial of risperidone augmentation for patients with difficult-to-treat unipolar, non-psychotic major depression. *J Psychiatr Res* 2009;**43**:205–14. <http://dx.doi.org/10.1016/j.jpsychires.2008.05.003>
70. Khullar A, Chokka P, Fullerton D, McKenna S, Blackman A. A double-blind, randomized, placebo-controlled study of quetiapine as augmentation therapy to SSRI/SNRI agents in the treatment of non-psychotic unipolar depression with residual symptoms, in *American Psychiatric Association 2006 Annual Meeting: New Research Abstracts (Toronto, Canada, May 20–25, 2006)*. Washington, DC: APA; 2006.
71. McIntyre A, Gendron A, McIntyre A. Quetiapine adjunct to selective serotonin reuptake inhibitors or venlafaxine in patients with major depression, comorbid anxiety, and residual depressive symptoms: a randomized, placebo-controlled pilot study. *Depress Anxiety* 2007;**24**:487–94. <http://dx.doi.org/10.1002/da.20275>
72. Earley W, McIntyre A, Bauer M, Pretorius HW, Shelton R, Lindgren P, *et al.* Efficacy and tolerability of extended release quetiapine fumarate (quetiapine extended release) as add-on to antidepressants in patients with major depressive disorder (MDD): results from a double-blind, randomized, phase III study. *American College of Neuropsychopharmacology 2007 Annual Meeting Abstracts (Boca Raton, FL, 9–13 December 2007)*. Nashville, TN: ACNP; 2007.
73. El-Khalili N, Joyce M, Atkinson S, Buynak R, Datto C, Lindgren P, *et al.* Adjunctive extended-release quetiapine fumarate (quetiapine-extended release) in patients with major depressive disorder and inadequate antidepressant response. *American Psychiatric Association 2008 Annual Meeting: New Research Abstracts (Washington, DC, 3–8 May 2008)*. Washington, DC: APA; 2008.
74. Heninger GR, Charney DS, Sternberg DE. Lithium carbonate augmentation of antidepressant treatment. An effective prescription for treatment-refractory depression. *Arch Gen Psychiatry* 1983;**40**:1335–42. <http://dx.doi.org/10.1001/archpsyc.1983.01790110077013>
75. Kantor D, McNevin S, Leichner P, Harper D, Krenn M. The benefit of lithium carbonate adjunct in refractory depression: fact or fiction? *Can J Psychiatry* 1986;**31**:416–18.
76. Zusky PM, Biederman J, Rosenbaum JF, Manschreck TC, Gross CC, Weilberg JB, *et al.* Adjunct low-dose lithium carbonate in treatment-resistant depression: a placebo-controlled study. *J Clin Psychopharmacol* 1988;**8**:120–4. <http://dx.doi.org/10.1097/00004714-198804000-00007>
77. Schöpf J, Baumann P, Lemarchand T, Rey M. Treatment of endogenous depressions resistant to tricyclic antidepressants or related drugs by lithium addition. Results of a placebo-controlled double-blind study. *Pharmacopsychiatry* 1989;**22**:183–7. <http://dx.doi.org/10.1055/s-2007-1014603>

78. Browne M, Lapiere YD, Hrdina PD, Horn E. Lithium as an adjunct in the treatment of major depression. *Int Clin Psychopharmacol* 1990;**5**:103–10. <http://dx.doi.org/10.1097/00004850-199004000-00004>
79. Joffe RT, Singer W, Levitt AJ, MacDonald C. A placebo-controlled comparison of lithium and triiodothyronine augmentation of tricyclic antidepressants in unipolar refractory depression. *Arch Gen Psychiatry* 1993;**50**:387–93. <http://dx.doi.org/10.1001/archpsyc.1993.01820170065008>
80. Stein G, Bernadt M. Lithium augmentation therapy in tricyclic-resistant depression. A controlled trial using lithium in low and normal doses. *Br J Psychiatry* 1993;**162**:634–40. <http://dx.doi.org/10.1192/bjp.162.5.634>
81. Nierenberg AA, Papakostas GI, Petersen T, Montoya HD, Worthington JJ, Tedlow J, et al. Lithium augmentation of nortriptyline for subjects resistant to multiple antidepressants. *J Clin Psychopharmacol* 2003;**23**:92–5. <http://dx.doi.org/10.1097/00004714-200302000-00013>
82. Baumann P, Nil R, Souche A, Montaldi S, Baettig D, Lambert S, et al. A double-blind, placebo-controlled study of citalopram with and without lithium in the treatment of therapy-resistant depressive patients: a clinical, pharmacokinetic, and pharmacogenetic investigation. *J Clin Psychopharmacol* 1996;**16**:307–14. <http://dx.doi.org/10.1097/00004714-199608000-00006>
83. Duru G, Fantino B. The clinical relevance of changes in the Montgomery–Åsberg Depression Rating Scale using the minimum clinically important difference approach. *Curr Med Res Opin* 2008;**24**:1329–35. <http://dx.doi.org/10.1185/030079908X291958>
84. Haynes RB, Wilczynski N, McKibbon KA, Walker CJ, Sinclair JC. Designing optimal search strategies for detecting clinically sound studies in MEDLINE. *J Am Med Assoc* 1994;**1**:447–59.
85. Dickersin K, Scherer R, Lefebvre C. Identifying relevant studies for systematic reviews. *BMJ* 1994;**309**:1286–91. <http://dx.doi.org/10.1136/bmj.309.6964.1286>
86. Benedict AA. Economic evaluation of duloxetine vs serotonin selective reuptake inhibitors and venlafaxine XR in treating major depressive disorder in Scotland. Cost-utility comparison of escitalopram and sertraline in the treatment of major depressive disorder. *J Affect Disord* 2010;**120**:94–104.
87. Xie F, Despiegel N, Danchenko N, Hansen K. Cost effectiveness analysis of escitalopram compared to venlafaxine and fluvoxamine in treatment of major depressive disorder. *Int J Psychiatry Clin Pract* 2009;**13**:59–69. <http://dx.doi.org/10.1080/13651500802450506>
88. Leelahanj T. The cost-effectiveness of aripiprazole as adjunctive therapy in major depressive disorder: Thai economic model. *J Med Assoc Thai* 2010;**93**(Suppl. 6):43–50.
89. Simpson KN, Welch MJ, Kozel FA, Demitrack MA, Nahas Z. Cost-effectiveness of transcranial magnetic stimulation in the treatment of major depression: a health economics analysis. *Adv Ther* 2009;**26**:346–68. <http://dx.doi.org/10.1007/s12325-009-0013-x>
90. Philips Z, Ginnelly L, Sculpher M, Claxton K, Golder S, Riemsma R, et al. Review of guidelines for good practice in decision-analytic modelling in health technology assessment. *Health Technol Assess* 2004;**8**(36). <http://dx.doi.org/10.2165/00019053-200624040-00006>
91. Revicki AD, Wood M. Patient-assigned health state utilities for depression-related outcomes: differences by depression severity and antidepressant medications. *J Affect Dis* 1998;**48**:25–36. [http://dx.doi.org/10.1016/S0165-0327\(97\)00117-1](http://dx.doi.org/10.1016/S0165-0327(97)00117-1)
92. Ministry of Health. *Clinical practice guidelines: depression*. Singapore: Ministry of Health; 2004.
93. Montgomery SA, Andersen HF. Escitalopram vs venlafaxine XR in the treatment of depression. *Int Clin Psychopharmacol* 2006;**21**:297–309.

94. Lancon C, Verpillat P, Annemans L, Despiegel N, Francois C. Escitalopram in major depressive disorder: clinical benefits and cost effectiveness vs citalopram. *Int J Psychiatry Clin Pract* 2007;**11**:44–52.
95. Posternak MA, Zimmerman M. Switching vs augmentation, a perspective naturalist comparison in depressed treatment resistant patients. *J Clin Psychiatry* 2001;**62**:135–42.
96. Adli M, Bschor T, Lucka C, Lewitza U, Ising M, Uhr M, *et al*. Long-term outcome after lithium augmentation in unipolar depression: focus on HPA system activity. *Neuropsychology* 2009;**60**:23–30. <http://dx.doi.org/10.1159/000234814>
97. Cegedim Strategic Data UK. *March 2005*. Psychotrak, HMSL Ltd, 2005.
98. House NA, Sommi RW. Defining response and attaining therapeutic goals for depression. *J Pharm Pract* 2001;**14**:453–7. <http://dx.doi.org/10.1177/089719001129040946>
99. Rush JA, Trivedi MH, Wisniewski SR, Nierenberg AA, Stewart JW, Warden D, *et al*. Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: a STAR*D report. *Am J Psychiatry* 2006;**163**:1905–17.
100. Gaynes BN, Rush AJ, Trivedi MH, Wisniewski SR, Spencer D, Fava M. The STAR*D study: treating depression in the real world. *Cleve Clin J Med* 2008;**75**:57–66.
101. Papaioannou D, Brazier JE, Paisley S. *NICE DSU technical support document 9. The identification, review and synthesis of health state utility values from the literature*. 2011. URL: www.nicedsu.org.uk (accessed 23 April 2013).
102. Bennett KJ, Torrance GW, Boyle MH, Guscott R. Cost-utility analysis in depression: the McSad utility measure for depression health states. *Psychiatr Serv* 2000;**51**:1171–6. <http://dx.doi.org/10.1176/appi.ps.51.9.1171>
103. Schaffer A, Levitt AJ, Hershkop SK, Oh P, MacDonald C, Lanctot K. Utility scores of symptom profiles in major depression. *Psychiatry Res* 2002;**110**:189–97. [http://dx.doi.org/10.1016/S0165-1781\(02\)00097-5](http://dx.doi.org/10.1016/S0165-1781(02)00097-5)
104. Pyne J, Sieber W, David K, Kaplan RM, Hyman Rapaport M, Keith Williams D. Use of the quality of well-being self-administered version (QWB-SA) in assessing health-related quality of life in depressed patients. *J Affect Disord* 2003;**76**:237–47. [http://dx.doi.org/10.1016/S0165-0327\(03\)00106-X](http://dx.doi.org/10.1016/S0165-0327(03)00106-X)
105. Sapin C, Fantino B, Nowicki ML, Kind P. Usefulness of EQ-5D in assessing health status in primary care patients with major depressive disorder. *Health Qual Life Outcomes* 2004;**2**:20.
106. Andresen E, Rothenberg R, Kaplan R. Performance of a self-administered mailed version of the Quality of Well-Being (QWB-SA) questionnaire among older adults. *Med Care* 1998;**36**:1349–60. <http://dx.doi.org/10.1097/00005650-199809000-00007>
107. Dolan P, Gudex C, Kind P, Williams A. The time trade-off method: results from a general population study. *Health Econ* 1996;**5**:141–54. [http://dx.doi.org/10.1002/\(SICI\)1099-1050\(199603\)5:2<141::AID-HEC189>3.0.CO;2-N](http://dx.doi.org/10.1002/(SICI)1099-1050(199603)5:2<141::AID-HEC189>3.0.CO;2-N)
108. National Institute for Health and Care Excellence (NICE). *Guide to the methods of technology appraisal*. June 2008. URL: www.nice.org.uk/media/B52/A7/TAMethodsGuideUpdatedJune2008.pdf (accessed April 2012).
109. Small GW, Birkett M, Meyers BS, Koran LM, Bystitsky A, Nemeroff CB. Impact of physical illness on quality of life and antidepressant response in geriatric major depression. *J Am Geriatr Soc* 1996;**4**:1220–5.

110. Dunner DL, Rush AJ, Russell JM, Burke M, Woodard S, Wingard P, *et al.* Prospective, long-term, multicenter study of the naturalistic outcomes of patients with treatment-resistant depression. *J Clin Psychiatry* 2006;**67**:688–95.
111. Bradley RH, Barkin RL, Jerome J, DeYoung K, Dodge CW. Efficacy of venlafaxine for the long term treatment of chronic pain with associated major depressive disorder. *Am J Ther* 2003;**10**:318–23. <http://dx.doi.org/10.1097/00045391-200309000-00003>
112. Doraiswamy PM, Khan ZM, Donahue RM, Richard NE. Quality of life in geriatric depression: a comparison of remitters, partial responders, and nonresponders. *Am J Geriatr Psychiatry* 2001;**9**:423–8. <http://dx.doi.org/10.1176/appi.ajgp.9.4.423>
113. Trivedi MH, Pigotti TA, Perera P, Dillingham KE, Carfango ML, Pitts CD. Effectiveness of low doses of paroxetine controlled release in the treatment of major depressive disorder. *J Clin Psychiatry* 2004;**65**:1356–64. <http://dx.doi.org/10.4088/JCP.v65n1010>
114. Trivedi MH, Dunner DL, Kornstein SG, Thase ME, Zajecka JM, Rothschild AJ, *et al.* Psychosocial outcomes in patients with recurrent major depressive disorder during 2 years of maintenance treatment with venlafaxine extended release. *J Affect Disord* 2010;**126**:420–9.
115. Carta MG, Hardoy MC, Pilu A, Sorba M, Floris AL, Mannu FA, *et al.* Improving physical quality of life with group physical activity in the adjunctive treatment of major depressive disorder. *Clin Pract Epidemiol Mental Health* 2008;**4**:1. <http://dx.doi.org/10.1186/1745-0179-4-1>
116. Karp JF, Skidmore E, Lotz M, Lenze E, Dew MA, Reynolds CF. Use of the late-life function and disability instrument to assess disability in major depression. *J Am Geriatr Soc* 2009;**57**:1612–19. <http://dx.doi.org/10.1111/j.1532-5415.2009.02398.x>
117. Fleck MPM. Efficacy of milnacipran in outpatients experiencing major depression non respondent to SSRIs: a 12-week open study. *Rev Psiquiatr Clin* 2010;**37**:241–50.
118. Barge-Schaapveld DQ, Nicolson NA, van der Hoop RG, De Vries MW. Changes in daily life experience associated with clinical improvement in depression. *J Affect Disord* 1995;**34**:139–54. [http://dx.doi.org/10.1016/0165-0327\(95\)00012-C](http://dx.doi.org/10.1016/0165-0327(95)00012-C)
119. Mitchell AJ, Selmes T. Why don't patients take their medicine? Reasons and solutions in psychiatry. *Adv Psychiatr Treat* 2007;**13**:336–46. <http://dx.doi.org/10.1192/apt.bp.106.003194>
120. Frank E, Prien RF, Jarret RB, Keller MB, Kupfer DJ, Lavori PW, *et al.* Conceptualization and rationale for consensus definitions of terms in major depressive disorder. *Arch Gen Psychiatry* 1991;**48**:851–5. <http://dx.doi.org/10.1001/archpsyc.1991.01810330075011>
121. Briggs AHE, Claxton K, Sculpher M. *Decision modelling for health economic evaluation*. Oxford: Oxford University Press; 2006.
122. Fekadu A, Wooderson SC, Rane LJ, Markopoulou K, Poon L, Cleare AJ. Long-term impact of residual symptoms in treatment-resistant depression. *Can J Psychiatry* 2011;**56**:549–57.
123. National Institute for Health and Care Excellence (NICE). *Bipolar disorder: the management of bipolar disorder in adults, children and adolescents, in primary and secondary care*. London: NICE; 2006.
124. Department of Health (DoH). *Reference costs 2010–11*. London. URL: www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_123459 (accessed April 2012).
125. Curtis L, Netten A. *Unit costs of health and social care 2011*. University of Kent, Canterbury: PSSRU; 2010.
126. Hospital Episode Statistics (HES) URL: www.hesonline.nhs.uk/Ease/servlet/ContentServer?siteID=1937&categoryID=202 (accessed 31 May 2012).

127. Meads CA, Cnossen JS, Meher S, Juarez-Garcia A, ter Riet G, Duley L, *et al.* Methods of prediction and prevention of pre-eclampsia: systematic reviews of accuracy and effectiveness literature with economic modelling. *Health Technol Assess* 2008;**12**(6).
128. Drug Patent Watch. *ABILIFY (aripiprazole)*. Otsuka Pharmaceuticals. URL: www.drugpatentwatch.com/ultimate/preview/tradename/index.php?query=ABILIFY (accessed August 2012).
129. Kroenke K, West SL, Swindle R, Gilsenan A, Eckert GJ, Dolor R, *et al.* Similar effectiveness of paroxetine, fluoxetine, and sertraline in primary care: a randomized trial. *JAMA* 2001;**286**:2947–55.

Appendix 1 Final protocol

HTA no. 10/301: Management of treatment resistant depression, PROTOCOL, June 2011

1. Title of the project

Lithium or an atypical antipsychotic in the management of treatment resistant depression: systematic review and economic evaluation

2. Name of TAR team and project 'lead'

BMJ Technology Assessment Group (BMJ-TAG), BMJ Evidence Centre, BMJ Group, London

Dr Steven J Edwards

Head of Health Technology Assessment

BMJ Technology Assessment Group (BMJ-TAG)

BMJ Evidence Centre

BMJ Group, BMA House

Tavistock Square

London WC1H 9JP

Tel: +44 (0) 207 383 6112

Mob: +44 (0) 776 823 7218

Fax: +44 (0) 207 383 6242

Email: SEdwards@BMJGroup.com

3. Plain English summary

Depression is a common mental disorder that presents with depressed mood, loss of interest or pleasure, feelings of guilt or low self-worth, disturbed sleep or appetite, low energy and poor concentration.¹

Depression can be categorised into two broad categories; unipolar depression and bipolar depression. People with unipolar depression suffer with only episodes of depression, whereas people with bipolar depression suffer with episodes of low mood, and abnormally elevated mood (also known as mania). The most common mood disorder is unipolar depression and because the pharmacological treatment of unipolar and bipolar depression are somewhat different we will be focusing on the people with unipolar depression in this report.

Depression may be treated with medication known as antidepressants, various kinds of psychological treatments or self-help measures.

There are lots of different antidepressant medications available and so if someone does not get better with their first treatment a different one may be tried.

This report will focus on people who have unipolar depression and who have not responded to treatment with at least two previous antidepressant medications; we refer to these people as having treatment resistant unipolar depression.

In people with treatment resistant depression it is thought that the addition of another medication such as lithium or an atypical antipsychotic drug could offer some benefit; however there is limited evidence directly comparing lithium and atypical antipsychotics in people with treatment resistant unipolar depression.²

The aim of this report is to identify how effective adding either lithium or an atypical antipsychotic medication to an antidepressant is at managing people with treatment resistant unipolar depression.

We also aim to perform an economic analysis to see how cost-effective these medications are when used to treat depression.

4. Decision problem

Background

Depression is a common mental disorder affecting about 121 million people worldwide and is among the leading causes of disability.¹ People presenting with depression may complain of depressed mood, loss of interest or pleasure, feelings of guilt or low self-worth, disturbed sleep or appetite, low energy and poor concentration.

Depression can be diagnosed clinically using different criteria, the most commonly used of which are the DSM IV criteria as published by the American Psychiatric Association and the ICD 10 criteria developed by the World Health Organisation.^{3,4}

Up to two thirds of patients with major depression will either not respond to or will have a sub-optimal response to first-line treatment with antidepressants (i.e. they may respond but not enter remission which is the relative absence of clinical symptomatology). There are several potential pharmacological treatment options for patients not achieving sufficient response with antidepressants, one of which is to augment the antidepressant with an agent not approved for use as monotherapy in major depressive disorder.⁵

Current NICE guidance² for the sequencing of treatments in depression after an inadequate response to at least one antidepressant recommends that people who are informed about and prepared to tolerate the increased side-effect burden, should be considered for treatment with the combination or augmentation of an antidepressant with lithium or an antipsychotic such as aripiprazole, olanzapine, quetiapine or risperidone or another antidepressant such as mirtazapine or mianserin.

Objective

This report aims to determine the clinical effectiveness and cost-effectiveness of SSRI antidepressant therapy with either lithium or an atypical antipsychotic in the management of people with treatment resistant unipolar depression.

For this review, treatment resistant depression will be defined as failure to respond to at least two previous antidepressant medications. We will not impose restrictions on the maximum number of previous antidepressant drugs allowed so as not to reduce the amount of data available for analysis as we aware that

there will be limited relevant SSRI RCT data available. This assumes that there is a consistent relative treatment effect independent of line of therapy, i.e. addition of an atypical or lithium has the same relative benefit whether given with third-line SSRI or fourth line SSRI, etc. However, a sensitivity analysis will be conducted to assess the impact of this assumption.

PICO criteria

The planned PICO is as follows:

- **Population:** Adults with treatment resistant unipolar depression defined as failure to respond to at least two previous antidepressants in the current episode of depression only.
- **Intervention:**
 - An **SSRI (selective serotonin reuptake inhibitor)** (defined as either Citalopram (Cipramil), Escitalopram (Cipralext), Fluoxetine (Prozac, Felicium, Prozep, Proxit), Fluvoxamine (Faverin), Paroxetine (Seroxat) or Sertraline (Lustral)), **PLUS**
 - An **atypical antipsychotic drug** (defined as either Amisulpride (Solian), Aripiprazole (Abilify), Clozapine (Clozaril, Denzapine, Zaponex), Olanzapine (Zyprexa, Zypadhera), Paliperidone (Invega), Quetiapine (Seroquel), Risperidone (Risperdal) or Ziprasidone (Geodon))
- **Comparator:**
 - An **SSRI** (defined as either Citalopram (Cipramil), Escitalopram (Cipralext), Fluoxetine (Prozac, Felicium, Prozep, Proxit), Fluvoxamine (Faverin), Paroxetine (Seroxat) or Sertraline (Lustral)) **PLUS**
 - **Lithium** (Lithium carbonate (Camcolit, Liskonum, Priadel) or Lithium citrate (Li-Liquid, Priadel) or Lithium (Litarex, Lithonate, Phasal))
- **Outcomes:**
 - Disease severity
 - Quality of life
 - Adverse effects
 - Withdrawals (all cause) as a surrogate outcome for adherence to medication
 - Relapse rate
 - Mortality
 - Cost-effectiveness

Subgroup analyses

The planned subgroup analyses are as follows:

- Different durations of depression (i.e. time since first onset of current episode of depression)
- Classes of previous antidepressants (e.g. SSRI or tricyclic antidepressant)
- Sex (i.e. males and females)
- Age (i.e. those < 75 years and those ≥ 75 years old)
- People with different severities of depression (i.e. based on trial entry Hamilton Depression Rating Scale rating)

Objectives

The key areas that we plan to address in this report are:

- To identify and review the existing evidence relating to the clinical outcomes as pre-specified above
- To report the cost-effectiveness of these treatments
- To identify what the potential areas for future research might be

5. Report methods for synthesis of evidence of clinical effectiveness

A review of the evidence for clinical effectiveness will be undertaken systematically following the general principles recommended in the PRISMA statement (formerly the QUOROM statement).⁶

Search strategy

The search strategy will comprise the following main elements:

1. Searching of electronic bibliographic databases
2. Contact with clinical experts in the field
3. Review of the reference lists of retrieved papers

1. The electronic databases that will be searched are EMBASE, MEDLINE, PsycINFO and the Cochrane Controlled Trials Register.

We will also search the ClinicalTrials.gov website to identify relevant ongoing clinical trials that when completed may have an impact on the results of this review, to assist us in drawing up our final recommendations.

2. We will contact clinical experts in the relevant therapy areas to request details of trials (published and unpublished) of which they may be aware. We will allow the experts 1 calendar month to provide an initial response, with any additional time allowed being dependent on whether we have reached the data analysis stage of the review.
3. The references from any relevant review papers or randomised controlled trials (RCTs) uncovered in the search will also be examined for additional references potentially relevant to the review.

Abstract appraisal

Titles and abstracts of studies identified by the search process will be assessed independently by two reviewers (VH and SB) for inclusion. In cases where the reviewers are unable to reach a consensus as to whether the full text should be obtained for further appraisal, the full text will be obtained.

When potentially relevant data are available in only an abstract format then we will attempt to contact the corresponding author in order to obtain the full publication; however, there will be a pre-specified deadline of 1 calendar month by which they will need to have contacted us, but we may allow additional time for them to supply the data requested depending on where we are in the review process. Any information supplied after the deadline will be included in only the discussion section of the review report.

Inclusion criteria

- For the review of clinical effectiveness, only RCTs will be included
- Adults \geq 18 years
- People with unipolar depression only
- Treatment resistant depression defined as failure to respond to at least two previous antidepressants in the current episode of depression only
- SSRI (selective serotonin reuptake inhibitor) given as baseline treatment and patient randomised to either lithium or an atypical antipsychotic
- Minimum duration of 4 weeks treatment with study medication for the current episode of depression
- Studies reporting on one or more of the following outcomes:
 - Disease severity
 - Quality of life

- Adverse effects
- Adherence to medication or withdrawals (all cause)
- Relapse rate
- Mortality
- Cost-effectiveness

Exclusion criteria

- Non-randomised studies
- Narrative reviews, editorials, opinions
- Studies performed in animals
- Studies not focusing on the treatment of the acute phase of depression (i.e. those only focusing solely on maintenance therapy)
- Bipolar depression or bipolar disorder diagnosis prior to study entry
- Underlying medical condition or another substantial co-morbid psychiatric condition
- Trials reporting only post-crossover results

Study inclusion assessment

Two reviewers (VH and SB) will independently assess for inclusion the full text of the trials identified during the abstract assessment stage and any differences in opinion will be arbitrated by a third reviewer (SJE).

Data extraction strategy

Data will be extracted by one reviewer (VH) using a standardised data extraction form (for draft copy of data collection form, please see appendix 10.2) and validated by second reviewer (SB).

A pragmatic decision for data validation will be made depending on the number of trials identified owing to the time constraints for completing this review. If a large number of trials are identified then all data will be validated (checked) by a second reviewer, with a sample being fully independently data extracted. This sample will be 25% or a minimum of 5 papers (whichever is larger).

The Data Extraction Form will be pilot tested on a sample of three papers by the reviewers and a final version agreed.

Discrepancies in the data extracted by the two reviewers will be resolved through discussion, with involvement of a third reviewer (SJE) if necessary.

Data from intention-to-treat (ITT) analyses will be extracted (per protocol (PP) data will also be extracted for use in a sensitivity analysis). Should a trial not report ITT data, we will treat missing data as treatment failures to allow our analysis to conform to an ITT analysis. For the purpose of this review, ITT will be defined as patients being analysed in the treatment group they were allocated to at randomisation regardless of whether they received the wrong intervention, withdrew or were lost to follow-up.

Study authors will be contacted to supply any additional information not included in published sources (including pre-crossover results in those trials reporting only post-crossover results) and there will be a pre-specified deadline by which we would require a response. The deadline will be 1 calendar month from the date of sending the request by which time they must have contacted us with at least an initial response acknowledging their intent to supply some of the information required. We may allow additional time for them to supply the data requested depending on where we are in the review process, however any information received after the deadline will be included in only the discussion section of the review.

Quality assessment strategy

Outcomes from the studies that meet the inclusion criteria will be assessed using the updated risk of bias tool developed by the Cochrane Collaboration (March 2011).⁷

These criteria assess the following areas:

1. Random sequence generation
2. Allocation concealment
3. Blinding of participants and personnel
4. Blinding of outcomes assessment
5. Incomplete outcome data
6. Selective reporting
7. 'Other bias'

Based on these criteria, an assessment for each outcome reported in the trial will be allocated based on the identified risk of bias. The three bias assessment categories used will be: low risk, high risk and unclear risk. Unclear risk is likely to be assigned owing to poor reporting of how the trial was conducted rather than a poorly conducted trial.⁸ Trials that are deemed to be at low or unclear risk of bias will be included in the main analysis; however, the trials rated high risk will be included in a sensitivity analysis.

Two reviewers (VH and SB) will independently rate the trial outcomes for inclusion and any differences in opinion will be arbitrated by a third reviewer (SJE). An outcome from an RCT will be considered appropriate for inclusion unless the trial demonstrates some feature that necessitates the exclusion of that outcome.

Methods of analysis/synthesis

Data will be tabulated and, where appropriate, meta-analysis will be employed to estimate a summary measure of effect on relevant outcomes based on ITT analyses (with a sensitivity analysis based on per protocol data).

We will not be assuming there is a class effect for any of the drugs included and so each individual drug will be considered separately in the review, i.e. each SSRI and atypical antipsychotic or lithium combination will form separate analyses.

Standard pairwise meta-analysis will be conducted when more than one trial is identified for inclusion for any pair of treatments under investigation. This will be carried out using a fixed effects model with the Mantel-Haenszel method.⁹ Sensitivity analysis will be conducted using a random effects model with the DerSimonian & Laird method.¹⁰

It is anticipated that a mixed treatment comparison (MTC; also called a multiple treatment meta-analysis and network meta-analysis) will need to be conducted to estimate the effects of the different treatments included in the research. A MTC can be seen as an extension of traditional pairwise meta-analysis.¹¹⁻¹³ It has advantages over standard pairwise meta-analysis as it is based on a network of connected trials where a new trial may enter the network if it is in a clinically comparable patient population, has a similar design to other trials incorporated in the network and contains at least one treatment that already exists within the network. It has been argued that this underlying assumption of exchangeability of data is no different from the practice within standard pairwise meta-analysis of combining similar trials.

The MTC will be conducted based on a fixed effects and a random effects model with the most appropriate model identified as the one with the lowest deviance information criterion (DIC).¹⁴ DIC measures the fit of the model while penalising for the number of effective parameters.^{12,15} For the chosen model, consistency of the evidence will be assessed using the posterior mean residual deviance, which should approximate the number of unconstrained data points in a good-fitting model.

For dichotomous outcomes we will use odds ratio as the summary statistic, and for continuous outcomes we will use the weighted mean difference as the summary statistic.

- Primary analysis will be:
 - Disease severity (measured by a reduction of at least 50% on Hamilton Depression Rating Scale (HDRS)¹⁶ or Montgomery-Åsberg Depression Rating Scale (MADRS).¹⁷ Where a study reports both we will use only the HDRS data).
- Secondary analyses will be:
 - Quality of life (QoL) as reported using a validated QoL rating scale¹⁸, e.g. EQ-5D, SF-36, HUI.
 - Adverse effects (data will be collected on those adverse effects most burdensome to patients such as agitation, akathisia, anxiety, cognitive dulling, constipation, diarrhoea, dry mouth, dyspepsia, extrapyramidal symptoms, kidney and thyroid dysfunction, lipid disturbance, gastrointestinal bleeding, (orthostatic) headache, hyperglycaemia, hypotension, nausea, polyuria, restlessness, sedation, sexual dysfunction, sleep disturbance, thirst, tremor, visual problems, and weight gain).
 - Withdrawals (all cause) as a surrogate outcome for adherence to medication
 - Relapse rate
 - Mortality (all cause)
- 8-week outcome data will be collected where reported. If 8-week data are not available, we will use outcome data reported from the nearest available time point
- Subgroup analyses will be performed in the following populations on only the primary outcome (disease severity), subject to the availability of data:
 - Different durations of depression (i.e. time since first onset of current episode of depression, short term < 6 months, long term > 6 months)
 - Classes of previous antidepressants (e.g. SSRI or tricyclic antidepressant)
 - Sex (i.e. males and females)
 - Age (i.e. those ≥ 75 years and those < 75 years old)
 - People with different severity's of depression, i.e. based on trial entry HDRS rating using the following categories: 16
 - 0–7 = Normal
 - 8–13 = Mild Depression
 - 14–18 = Moderate Depression
 - 19–22 = Severe Depression
 - ≥ 23 = Very Severe Depression

In the absence of suitable data to perform a meta-analysis, the available data will be tabulated where possible and discussed in a narrative review.

Heterogeneity

In addition to the existing pre-specified subgroups, other potential sources of clinical heterogeneity could be a result of combining different preparations of drugs.

For pairwise meta-analysis, heterogeneity will be explored through consideration of the study populations, methods and interventions, by visualisation of results and, in statistical terms, by the χ^2 test for homogeneity and the I^2 statistic. Statistically significant heterogeneity will be defined as $p < 0.10$. Levels of inconsistency will be assessed using I^2 and will be defined as follows: I^2 of: 0%–25% = low level of inconsistency; 26%–50% = moderate level of inconsistency; and > 50% = high level of inconsistency.¹⁹

If statistically significant heterogeneity is detected in any of the primary or secondary analyses, hypothesis-generating subgroup analysis will be conducted, but the results from such analyses will be treated with caution. Meta-regression will be attempted if significant statistical heterogeneity is identified among trials analysed and there are 10 or more trials in the review.

For the MTC, where a random effects model is deemed the best fit, the degree of heterogeneity will be investigated by evaluating the posterior mean tau-squared. Where possible, any closed loops formed by the network of trials will be assessed separately to determine if the results from the 'direct' evidence is coherent with the 'indirect' evidence when the wider network is introduced. Any incoherence identified will be investigated.

Sensitivity analysis

Sensitivity analyses are planned on the primary analysis and consist of:

- different number of prior antidepressants for the current episode of depression;
- assuming a 'class' effect with SSRIs and atypical antipsychotics;
- changing the quality assessment to include the trial outcomes excluded on grounds of methodological quality; i.e. those categorised as of high risk of bias;
- changing the analysis from using ITT (intention to treat) data to per protocol data.

Publication bias

For each of the primary pairwise meta-analyses, a funnel plot will be used to assess publication bias. A regression of normalised effect compared with precision will also be calculated as a test for small study effects (using a $p < 0.10$ as an indicator of a significant result).²⁰

6. Report methods for synthesising evidence of cost-effectiveness

Identifying and systematically reviewing published cost-effectiveness studies

The following databases will be used to identify studies of the cost-effectiveness. MEDLINE, EMBASE, PsycINFO, CINAHL NHS Economic Evaluation Database, Health Technology Assessment Database and Office of Health Economics Health Economic evaluation database. We will apply a cost search filter to the comprehensive clinical search strategy described in *Section 5*.

In order to express clinical outcomes in the form of QALYs, utility weights for health states relating to treatment resistant depression are required. Utility weights represent the health related quality of life (HRQOL) associated with specific health states; they are estimated based on people's preferences and perceptions of quality of life characterising the health states under consideration. We will undertake a systematic quality of life search where health economics and quality-of-life search filters will be used in MEDLINE, EMBASE, PsycINFO and CINAHL to identify relevant studies.

The inclusion and exclusion criteria for economic evaluations will be the same as those for the systematic review of clinical effectiveness and in addition the health economic evaluation will also include:

- non-randomised studies will be included (e.g. decision-model based analysis or analysis of person-level cost and effectiveness data alongside observational studies)
- full cost-effectiveness analyses, cost–utility analyses, cost–benefit analyses and cost–consequence analyses will be included
- stand-alone UK cost analysis will also be sought and appraised

Titles and abstracts returned by the search strategy will be assessed independently by two health economists (LN and NT) and screened for possible inclusion. Any disagreements will be resolved by a third health economist (SJE).

Evaluation of costs and cost-effectiveness (may include development of a *de novo* economic model)

The methodological quality of economic evaluations will be assessed according to internationally accepted criteria such as the Consensus on Health Economic Criteria list questions developed by Evers *et al.* (2005).²¹ Any studies based on decision models will be assessed using the checklist developed by Phillips *et al.* (2004).¹⁸

In addition, a new economic evaluation will be carried out from the perspective of the UK NHS using a probabilistic decision-analytic (Markov) modelling approach to estimate the costs and QALYs of SSRI with an atypical antipsychotic compared with SSRI with lithium in the management of treatment resistant unipolar depression. An annual discount rate of 3.5% will be used for both costs and QALYs in accordance with NICE guidance.²² Model structure, data inputs and modelling assumptions will be determined in consultation with clinical experts to ensure they reflect the best current clinical practice and evidence. Uncertainty in the data used to populate the model will be characterised using appropriate methods, such as probabilistic sensitivity analysis. The time horizon of our analysis will preferably be a patient's lifetime in order to reflect the chronic nature of the disease. However time horizon may be dictated by the availability of data in which case shorter time horizons will be modelled.

Ideally, evidence on the impact of these therapies on HRQoL will be available directly from the trials included within the review. In the absence of such evidence, the mathematical model may use indirect evidence on quality of life from alternative sources, such as related technology appraisals or clinical guidelines. Quality of life data will be reviewed and used to generate the quality adjustment weights required for the model. We will also adjust utility for age using data from the Health Survey of England.²³

Results will be presented as incremental cost-effectiveness ratios (ideally cost per quality adjusted life year) and cost-effectiveness acceptability curves, which quantify the degree of uncertainty.

7. Expertise in this TAR team

TAR Centre

The BMJ Evidence Centre comprises over 30 specialists with a wealth of experience in diverse health-related areas and includes clinicians, pharmacists, information specialists, health informatics specialists, project managers, systematic reviewers, clinical guideline developers and health economists.

The BMJ-TAG core team consists of 5 members. Together, we have an array of experience amongst us in producing focussed reports in a short timescale for policy customers such as NICE. Please see below for further details of each team member's experience.

- Dr Steven J Edwards DPhil MSc BSc (Hons), Head of Health Technology Assessment:** Over the past 12 years, Steve has conducted over 40 systematic reviews and health economic evaluations in a range of therapeutic areas including cardiovascular, CNS, gastroenterology, infection, oncology and respiratory medicine. His interests are in the use of the best available evidence for decision making with an emphasis on the design and conduct of clinical trials, systematic reviews, meta-analyses, adjusted indirect comparisons and their subsequent use in economic evaluations. His postgraduate research in this area at the University of Oxford resulted in him being awarded the first doctorate of evidence based health care. In addition, Steve is an honorary senior lecturer in health economics at the London School of Hygiene & Tropical Medicine, a member of the Cochrane Statistical Methods Group, the Campbell &

Cochrane Economics Methods Group, and an Editorial Board member of the International Journal of Clinical Practice.

- **Dr Samantha Barton PhD BSc (Hons), Health Technology Assessment Analyst:** Sam has extensive experience in the critical appraisal of studies. During the past 4 years, she has contributed to the publication of over 50 systematic reviews on prevention and treatment of various clinical conditions. She has worked on reviews in the areas of mental health, sexual health, infectious diseases, cardiovascular disorders, respiratory disorders and oncology.
- **Dr Victoria Hamilton MBChB, Health Technology Assessment Analyst:** Vicky has a clinical background with relevant experience in the fields of general surgery, general medicine, general practice, paediatrics and orthopaedic surgery. Vicky also has experience in the critical appraisal of clinical studies and over the last year has contributed to the publication of systematic reviews in a variety of clinical areas. She also has experience in the process and use of clinical audit to review current clinical practice within both primary and secondary care settings.
- **Mr Leo Nherera MSc BSc (Hons), Health Economist:** Over the past 6 years, Leo has been working for the NICE clinical guideline programme and has successfully worked in eight published clinical guidelines and one Public Health guideline. His work involved appraising economic evaluations as well as doing original economic analysis for various guideline questions to assist in guideline recommendations. Leo was involved in organising and teaching the Health Economics module at Queen Mary University of London. He has also peer-reviewed papers for the International Journal of Clinical Practice. His interests are in the use of the best available evidence for decision making with an emphasis on systematic reviews and meta-analyses and their subsequent use in economic evaluations.
- **Ms Nicola Trevor MSc BSc (Hons), Health Economist:** Nicola has a strong mathematical background, with a Masters in analytical, numerical and statistical modelling techniques, which over the past 2 years she has applied in the field of health economics, conducting economic evaluations and statistical analysis for systematic review in disease areas such as multiple sclerosis, cardiovascular disease, Gaucher's disease and oncology. Her interests are in the use of the best available techniques for decision making with an emphasis on survival analysis, meta-analysis, modelling approaches and the use of Bayesian methods in economic evaluations.

Recent publications from the team members include:

Nherera L, Marks D, Minhas R, *et al.* Probabilistic cost-effectiveness analysis of cascade screening for Familial Hypercholesterolaemia using alternative diagnostic and identification strategies. *Heart* 2011;**97**:1175–81.

National Collaborating Centre for Women's and Children Health. Multiple pregnancy: The management of twin and triplet pregnancy in the antenatal period London, Royal College of Obstetricians and Gynaecologists, 2011; (in press).

Edwards SJ, Wordsworth S, Clarke MJ. Treating pneumonia in critical care in the UK following failure of initial antibiotic: a cost-effectiveness analysis comparing meropenem with piperacillin/tazobactam. *European Journal of Health Economics* 2011;**12**: (available online first at: www.springerlink.com/content/q044j5t32601vt4l/).

Halpin DMG, Gray J, Edwards SJ, *et al.* Budesonide/formoterol versus salmeterol/fluticasone in COPD: a systematic review and adjusted indirect comparison of pneumonia in randomized controlled trials. *International Journal of Clinical Practice* 2011;**65**:764–74.

Edwards SJ, Borrill J. Network meta-analysis: importance of appropriate trial selection. *Value in Health* 2010; **13**:681–2.

Edwards SJ, von Maltzahn R, Naya IP, *et al.* Budesonide/formoterol for maintenance and reliever therapy: a meta-analysis of randomised controlled trials. *International Journal of Clinical Practice* 2010; **64**:619–27.

Gray J, Edwards SJ, Lip GYH. Comparison of sequential rosuvastatin doses in hypercholesterolaemia: a meta-analysis of randomized controlled trials. *Current Medical Research and Opinion* 2010;**26**:537–47.

Trevor NC, Alnwick K. How can the use of predictive biomarkers lead to positive HTA recommendations? *Value in Health* 2010;**13**:A423–4.

Trevor NC, Tang M, Samuels ER. Investigating the impact of R&D investment and policy on innovative performance in Europe. *Value in Health* 2010;**13**:A414.

Edwards SJ, Gray J. Budesonide/formoterol plus tiotropium (BUD/FORM+TIO) vs salmeterol/fluticasone plus tiotropium (SALM/FLU+TIO): a systematic review and adjusted indirect comparison between two alternative triple treatments in chronic obstructive pulmonary disease (COPD). *Value in Health* 2010;**13**:A319.

Edwards SJ, Welton NJ, Borrill J. Gefitinib compared with doublet chemotherapy for first-line treatment non-small-cell lung cancer (NSCLC) a systematic review and adjusted indirect comparison. *Value in Health* 2010;**13**:A252–3.

Edwards SJ, Welton NJ, Borrill J. Tolerability of first-line treatments of locally advanced or metastatic non-small-cell lung cancer (NSCLC) a systematic review and adjusted indirect comparison. *Value in Health* 2010;**13**:A250.

Nherera L, Calvert NW, DeMott K, *et al.* Cost effectiveness analysis of the use of a high intensity statin compared to a low intensity statin in the management of patients with familial hypercholesterolaemia. *Current Medical Research and Opinion* 2010;**26**:529–36.

Visintin C, Mugglestone MA, Almerie MQ, *et al.* on behalf of the Guideline Development Group. Management of hypertensive disorders during pregnancy: summary of NICE guidance. *British Medical Journal* 2010;**341**:c2207.

National Collaborating Centre for Women's and Children Health. Hypertension in pregnancy: The management of hypertensive disorders during pregnancy. London, Royal College of Obstetricians and Gynaecologists 2010.

External Clinical Expert Advisors

Professor Philip J. Cowen – MRC Clinical Scientist and Professor of Psychopharmacology; Specialist in Psychopharmacology of Mood Disorders

Neurosciences Building,

Warneford Hospital,

Oxford OX3 7JX,

United Kingdom

phil.cowen@psych.ox.ac.uk

Recent publications include:

- McCabe C, Mishor Z, Cowen PJ, *et al.* Diminished neural processing of aversive and rewarding stimuli during selective serotonin reuptake inhibitor treatment. *Biological Psychiatry* 2010;**67**:439–45.

- Harmer CJ, O'Sullivan U, Favaron E, *et al.* Effect of acute antidepressant administration on negative affective bias in depressed patients. *American Journal of Psychiatry* 2009;**166**:1178–84.
- Harmer CJ, Goodwin GM, Cowen PJ. Why do antidepressants take so long to work? A cognitive neuropsychological model of antidepressant drug action. *British Journal of Psychiatry* 2008;**195**:102–8.
- Gelder M, Cowen P, Harrison P. *Shorter Oxford Textbook of Psychiatry*. Oxford University Press, Oxford, 1995, 2001, 2006.

Dr Luiz Dratcu – Consultant Psychiatrist and Specialist in Psychopharmacology, Treatment Resistant Mental Illness, Schizophrenia and Affective Disorders

Maudsley Hospital

South London & Maudsley NHS Foundation Trust

Denmark Hill,

London SE5 8AZ,

United Kingdom

luiz.dratcu@slam.nhs.uk

Recent publications include:

- Dratcu L. The quest for the pharmacological treatment of schizophrenia: from conventional neuroleptics to atypical anti-psychotics and beyond. *Vertex* 2010;**21**:385–93.
- Dratcu L. The future of depression: a complex neuroendocrine, inflammatory and neurodegenerative systemic illness. *Vertex* 2009;**20**:329–41.
- Dratcu L, Grandison A, McKay G, *et al.* Clozapine-resistant psychosis, smoking, and caffeine: managing the neglected effects of substances that our patients consume every day. *American Journal of Therapeutics* 2007;**14**:314–8.
- Dratcu L, Olowu P, Hawramy M, *et al.* Aripiprazole in the acute treatment of male patients with schizophrenia: effectiveness, acceptability, and risks in the inner-city hospital setting. *Neuropsychiatric Disease and Treatment* 2006;**2**:191–7.

8. Competing interests of authors

Steve Edwards has previously been an employee of AstraZeneca, which holds the marketing authorisation for Seroquel® (quetiapine). He has no ongoing financial connection nor owns significant shares with AstraZeneca.

Professor Philip J. Cowen has received consultancy fees from Servier, Lundbeck and Eli Lilly, and fees for speaking from AstraZeneca, Servier and Lundbeck. He has also provided advice to legal representatives of GSK.

Dr Luiz Dratcu has received consultancy fees, fees for speaking and hospitality from BMS/Otsuka and Merck. He has also received hospitality from Lilly.

9. Timetable/milestones

Finalise protocol – June 2011

Send progress report to NETSCC, HTA – February 2012

Submit assessment report to NETSCC, HTA – March 2012

The timetable is based on a 6-month working time-frame, commencing in mid-July assuming that the final approval of the protocol has been received by this time.

Timelines may be subject to change in the event of any additional urgent work commitments such as STA work for NICE; however we will endeavour to inform NETSCC of any commitments which may delay the completion of this project at the earliest possible date.

10. Appendices

10.1.1. Draft MEDLINE search strategy (Clinical)

10.1.2. Draft MEDLINE search strategy (Health Economics and Quality of life)

10.2. Data extraction form

10.3 Team members' contributions

10.4 References

Appendix 10.1.1 Draft MEDLINE search strategy

< 1948 to June Week 1 2011 >

Search Strategy:

1. Randomized Controlled Trials as Topic/ (73,451)
2. randomized controlled trial/ (308,386)
3. Random Allocation/ (71,692)
4. Double Blind Method/ (110,600)
5. Single Blind Method/ (15,044)
6. clinical trial/ (463,236)
7. clinical trial, phase i.pt. (11,244)
8. clinical trial, phase ii.pt. (17,834)
9. clinical trial, phase iii.pt. (6176)
10. clinical trial, phase iv.pt. (614)
11. controlled clinical trial.pt. (82,578)
12. randomized controlled trial.pt. (308,386)
13. multicenter study.pt. (131,287)
14. clinical trial.pt. (463,236)
15. exp Clinical Trials as topic/ (242,013)
16. or/1-15 (859,148)
17. (clinical adj trial\$.tw. (155,138)
18. ((singl\$ or doubl\$ or treb\$ or tripl\$) adj (blind\$3 or mask\$3)).tw. (107,934)
19. PLACEBOS/ (29,733)
20. placebo\$.tw. (129,547)

21. randomly allocated.tw. (12,594)
22. (allocated adj2 random\$.tw. (14,831)
23. or/17-22 (326,697)
24. 16 or 23 (954,423)
25. case report.tw. (158,367)
26. letter/ (716,157)
27. historical article/ (275,084)
28. or/25-27 (1,139,766)
29. 24 not 28 (928,335)
30. exp Depression/ or exp Depressive Disorder/ (127,115)
31. (depress* or adjustment disorder* or mood disorder* or affective disorder* or affective symptom* or dysthymi* or dysphori*).mp. (344,159)
32. 30 or 31 (344,159)
33. 29 and 32 (39,592)
34. Serotonin Uptake Inhibitors/ or Antidepressive Agents, Second-Generation/ or ssri*.mp. or exp Serotonin
35. Antagonists/ (59,672)
36. citalopram.mp. or exp Citalopram/ (4058)
37. escitalopram.mp. (779)
38. fluoxetine.mp. or exp Fluoxetine/ (9218)
39. fluvoxamine.mp. or exp Fluvoxamine/ (2286)
40. paroxetine.mp. or exp Paroxetine/ (4469)
41. sertraline.mp. or exp Sertraline/ (2966)
42. 36 or 37 or 38 or 39 or 40 (16,313)
43. 35 and 41 (1722)
44. 35 or 37 or 38 or 39 or 40 (18,583)
45. 36 and 43 (713)
46. 35 or 36 or 38 or 39 or 40 (11,526)
47. 37 and 45 (2095)
48. 35 or 36 or 37 or 39 or 40 (17,294)
49. 38 and 47 (931)
50. 35 or 36 or 37 or 38 or 40 (15,845)
51. 39 and 49 (1665)
52. 35 or 36 or 37 or 38 or 39 (16,994)
53. 40 and 51 (1311)
54. 42 or 44 or 46 or 48 or 50 or 52 (3310)
55. 33 and 53 (799)
56. lithium.mp. or exp Lithium Carbonate/ or exp Lithium/ or exp Lithium Compounds/ or exp Lithium Chloride/ (29,746)
57. (antipsychotic* or anti?psychotic* or anti-psychotic*).mp. (41,739)
58. amisulpride.mp. (571)
59. aripiprazole.mp. (1454)
60. clozapine.mp. (8530)
61. olanzapine.mp. (5275)
62. paliperidone.mp. (153)
63. quetiapine.mp. (2428)
64. risperidone.mp. (5910)
65. or/34-40 (66,927)
66. or/55-63 (72,543)
67. 33 and 64 and 65 (713)
68. 54 or 66 (1455)

10.1.2 *Draft MEDLINE search strategy (Health Economics and Quality of life)*

Economics search terms

1. exp economics/ (438,053)
2. exp Costs and Cost Analysis/ (38,816)
3. Cost Benefit Analysis/ (51,007)
4. value of life/ (5162)
5. exp models economic/ (7945)
6. exp fees/and charges/ (7703)
7. exp budgets/ (10,939)
8. (economic adj2 burden).tw. (2622)
9. (expenditure* not energy).tw. (14,210)
10. budget*.tw. (14,415)
11. (economic* or price* or pricing or financ* or fee* or pharmacoeconomic* or pharmaeconomic* or pharmaco-economic*).tw. (128,436)
12. (decision adj1 (tree* or analys* or model*)).tw. (6411)
13. Resource Allocation/ (6522)
14. (unit cost or unit-cost or unit-costs or unit costs or drug cost or drug costs or hospital costs or health-care costs or health care cost or medical cost or medical costs).tw. (16,355)
15. ((value or values or valuation) adj2 (money or monetary or life or lives or costs or cost)).tw. (3225)
16. Markov Chains/ (7220)
17. exp Decision Support Techniques/ (48,239)
18. (resource adj2 (use* or utili* or allocat*)).tw. (10,801)
19. (cost adj2 (util* or effective* or efficac* or benefit* or consequence* or analys* or minimi* or allocation* or control* or illness* or affordable* or fee* or charge* or charges)).tw. (71,017)
20. or/1-19 (627,358)

*Combining condition, intervention, comparator and cost terms gets a total of **36** studies potential cost-effectiveness abstracts.*

Quality of life search terms

1. exp quality of life/ (90,943)
2. quality of life.tw (100,676)
3. life quality.tw (2525)
4. (sf 36 or sf36 or sf thirty six or sf thirty six or short form 36 or short form thirty six or short form thirtysix or shortform 36).tw (11,072)
5. (euroqol or eq5d or eq 5d).tw (2147)
6. quality adjusted life\$.tw (3963)
7. (QALY\$ or lifeyear\$ or life year\$ or ((qualit\$3 or value) adj3 (life or survival))).tw. (108,136)
8. ((burden adj3 (disease or illness)) or (resource adj3 (allocation\$ or utilit\$)) or (value adj5 money)).tw. (12,216)
9. (budget\$ or cost\$ or econom\$ or expenditure\$ or financ\$ or fiscal\$ or funding or pharmacoeconomic\$ or price or prices or pricing).tw. (441,366)
10. (Hamilton depression rating scale\$.ab. (2004)
11. (Montgomery-Åsberg depression rating scale\$.ab. (1004)
12. or/1-11 (575,570)

*Combining condition, intervention, comparator and quality of life terms gets a total of **136** potential quality of life abstracts.*

Appendix 10.2 Data extraction form

PART ONE: REVIEW, REVIEWER AND STUDY INFORMATION

Study ID:

Reviewer name:

Date of completion of this form:

Title of paper/abstract:

Source (journal, year, volume, pages):

Authors:

Language of publication:

Type of paper (e.g. full paper/abstract/poster):

PART TWO: VERIFICATION OF STUDY ELIGIBILITY

Type of clinical trial

1) Is the study randomised? YES UNCLEAR NO

Population in the clinical trial

2) Is the population adults ≥ 18 years old? YES UNCLEAR NO

3) Did the RCT include people with unipolar depression? YES UNCLEAR NO

4) Did the RCT include people with treatment resistant depression (defined as failure to respond to ≥ 2 antidepressants)? YES UNCLEAR NO

Interventions in the clinical trial

5) Does the trial compare SSRI + atypical antipsychotic or lithium or no treatment with SSRI + lithium or atypical antipsychotic or placebo or no treatment? YES UNCLEAR NO

6) Did both groups experience the same care except for the two interventions under investigation? YES UNCLEAR NO

Outcomes of the clinical trial

7) Does the study report on outcomes during the treatment of the acute phase of depression? YES UNCLEAR NO

8) Did the clinical trial investigate at least one of the following: disease severity, quality of life, adverse events, withdrawals (all cause), relapse rate, mortality (all cause)? YES UNCLEAR NO

9) Are the outcomes measured after ≥ 4 weeks treatment with study medication? YES UNCLEAR NO

If you answered NO to any of the above questions do not proceed to Part 3.

PART THREE: INFORMATION ABOUT THE STUDY

Characteristics of the trial

Country(ies) where the clinical trial was conducted:

Sponsors of the clinical trial:

Any conflicts of interest reported for any of the researchers?

Date the clinical trial was conducted:

Type of clinical trial design (e.g. parallel, crossover, or cluster trial):

If the trial was of crossover design, are there pre-crossover results reported?

Was the trial multicentre? If so, how many centres were there?

Characteristics of the patients

Inclusion criteria: how and where were patients enrolled, were any patient risk factors used? What details of the antidepressant(s) patients had failed to respond to are provided?

Exclusion criteria: were specific groups of people excluded?

Total number of people randomised:

Information on the age of the patients:

Information on the sex of the patients (m/f):

Information on the ethnicity of the patients:

Information on patients' medical history (i.e. previous depression):

Type of intervention

Intervention 1: SSRI + XX (where XX = atypical antipsychotic or lithium or no treatment)

SSRI name and brand:

SSRI dose and regimen used (e.g. 80 mg OD):

Delivery of SSRI (e.g. PO tablet/dissolvable/enteric coated):

Number of doses of SSRI given per day (with SD/SE if given):

Duration of SSRI treatment in days (with SD/SE if given):

What was XX (name and brand)?

XX dose and regimen used (e.g. 80 mg OD):

Delivery of XX (e.g. PO tablet/dissolvable/enteric coated):

Number of doses of XX given per day (with SD/SE if given):

Duration of XX treatment in days (with SD/SE if given):

Number of patients randomised:

Intervention 2: SSRI + YY (where YY = lithium or atypical antipsychotic or placebo or no treatment)

SSRI name and brand:

SSRI dose and regimen used (e.g. 80 mg OD):

Delivery of SSRI (e.g. PO tablet/dissolvable/enteric coated):

Number of doses of SSRI given per day (with SD/SE if given):

Duration of SSRI treatment in days (with SD/SE if given):

What was YY (name and brand)?

YY dose and regimen used (e.g. 80 mg OD):

Delivery of YY (e.g. PO tablet/dissolvable/enteric coated):

Number of doses of YY given per day (with SD/SE if given):

Duration of YY treatment in days (with SD/SE if given):

Number of patients randomised:

Was the formulation and appearance of YY (e.g. lithium) matched to that of XX (e.g. atypical antipsychotic)?

Were any additional interventions given to either or both groups?

Types of outcome

Which of the following outcomes have been assessed in the clinical trial?	YES	UNCLEAR	NO
Disease severity?	YES	UNCLEAR	NO
How was disease severity defined in the clinical trial?	YES	UNCLEAR	NO
Quality of life?	YES	UNCLEAR	NO
How was quality of life defined in the clinical trial?	YES	UNCLEAR	NO
Adverse events?	YES	UNCLEAR	NO
How were adverse events defined in the clinical trial? (e.g. investigator attributed?)	YES	UNCLEAR	NO
Withdrawal (all cause)?	YES	UNCLEAR	NO
How was withdrawal defined in the clinical trial?	YES	UNCLEAR	NO
Relapse rate?	YES	UNCLEAR	NO
How was relapse rate defined in the clinical trial?	YES	UNCLEAR	NO

All-cause mortality?	YES	UNCLEAR	NO
How was all-cause mortality defined in the clinical trial?	YES	UNCLEAR	NO
Any other outcomes reported in trial (please list)?	YES	UNCLEAR	NO

ITT data collection table:

Outcomes		Timeframe (weeks)	SSRI + XX		SSRI + YY	
			<i>n</i>	<i>N</i>	<i>n</i>	<i>N</i>
Disease severity	50% reduction in HDRS					
	50% reduction in MADRS					
Quality of life	Trial scale:					
Withdrawals (all cause)						
Relapse rate						
All-cause mortality						
Adverse events (please specify)						

n = number of patients with the outcome; *N* = number of patients assessed

Per protocol data collection table:

Outcomes		Timeframe (weeks)	SSRI + XX		SSRI + YY	
			<i>n</i>	<i>N</i>	<i>n</i>	<i>N</i>
Disease severity	50% reduction in HDRS					
	50% reduction in MADRS					
Quality of life	Trial scale:					
Withdrawals (all cause)						
Relapse rate						
All-cause mortality						
Adverse events (please specify)						

n = number of patients with the outcome; *N* = number of patients assessed

Did the RCT carry out any subgroup analyses of interest? (i.e. Different durations of depression, different classes of previous antidepressants, different genders, age, different severity's of depression or different number of prior antidepressants)

If yes, please give details here.

PART FOUR: CLINICAL TRIAL QUALITY

Please describe the method of randomisation and allocation concealment used in the clinical trial:

Please describe the method of blinding and who was blinded in the clinical trial:

Please describe the number of patients lost to follow up (the overall number and number by treatment group, give reasons for loss to follow up):

How would you describe the trials design to minimise bias for (please tick):

Outcome	Risk of bias	Low risk	Unclear risk	High risk	Comments
Disease severity	Random sequence generation				
	Allocation concealment				
	Blinding (participants & personnel)				
	Blinding of outcomes assessment				
	Incomplete outcome data				
	Selective reporting				
	'Other bias'				
Quality of life	Random sequence generation				
	Allocation concealment				
	Blinding (participants & personnel)				
	Blinding of outcomes assessment				
	Incomplete outcome data				
	Selective reporting				
	'Other bias'				
Withdrawals (all cause)	Random sequence generation				
	Allocation concealment				
	Blinding (participants & personnel)				
	Blinding of outcomes assessment				
	Incomplete outcome data				
	Selective reporting				
	'Other bias'				
Relapse rate	Random sequence generation				
	Allocation concealment				
	Blinding (participants & personnel)				
	Blinding of outcomes assessment				
	Incomplete outcome data				
	Selective reporting				
	'Other bias'				
All-cause mortality	Random sequence generation				
	Allocation concealment				
	Blinding (participants & personnel)				
	Blinding of outcomes assessment				
	Incomplete outcome data				
	Selective reporting				
	'Other bias'				

Outcome	Risk of bias	Low risk	Unclear risk	High risk	Comments
Adverse events	Random sequence generation				
	Allocation concealment				
	Blinding (participants & personnel)				
	Blinding of outcomes assessment				
	Incomplete outcome data				
	Selective reporting				
	'Other bias'				

How would you rate the trials overall risk of bias? Low risk Unclear High risk

Do you have any additional comments you would like to make about this clinical trial?

Ideally, would you like further information about the clinical trial from the authors (If so, please give details)?

Appendix 10.3 Team members' contributions

Steve Edwards, Head of HTA, will develop the protocol, act as the third reviewer for assessment of trials and cost-effectiveness studies, validate data extraction and any data analysis required, validate the economic model, contribute to writing/editing of the report, be overall director of the project and act as guarantor of the report.

Sam Barton, HTA Analyst, will act as co-reviewer for assessing trials for inclusion and data extraction, and contribute to the writing/editing of the report.

Vicky Hamilton, HTA Analyst, will provide overall project management, develop the protocol, write and run the search strategy, act as co-reviewer for assessing trials for inclusion and data extraction (and perform data analysis as required), and contribute to the writing/editing of the report.

Leo Nherera, Health Economist, will develop the protocol, act as co-reviewer of the cost-effectiveness studies, develop the economic model, and contribute to the writing/editing of the report.

Nicola Trevor, Health Economist, will act as co-reviewer of the cost-effectiveness studies, validate the economic model, and contribute to the writing/editing of the report.

Professor Cowen and Dr Dratcu, Clinical Expert Advisors, will provide clinical advice as required throughout the protocol development and review processes.

Appendix 10.4 References

1. http://www.who.int/mental_health/management/depression/definition/en/
2. National Institute for Health and Clinical Excellence (2009) *Depression: the treatment and management of depression in adults* (update). CG90. London: National Institute for Health and Clinical Excellence.
3. *DSM-IV-TR Diagnostic and Statistical Manual of Mental Disorders Fourth Edition, Text Revision*. DOI: 10.1176/appi.books.9780890423349
4. <http://www.who.int/classifications/icd/en/>
5. Nelson JC, Papakostas GI. Atypical Antipsychotic Augmentation in Major Depressive Disorder: A Meta-Analysis of Placebo-Controlled Randomized Trials. *Am J Psychiatry*. 2009;**166**:980–991.

6. Schulz KF, Altman DG, Moher D, for the CONSORT Group. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. *Ann Int Med*. 2010;**152**. Epub 24 March.
7. Higgins JPT, Green S (editors). *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available from www.cochrane-handbook.org
8. Soares HP, Daniels S, Kumar A, Clarke M, Scott C, Swann S, Djulbegovic B. Radiation Therapy Oncology Group. Bad reporting does not mean bad methods for randomised trials: observational study of randomised controlled trials performed by the Radiation Therapy Oncology Group. *BMJ*. 2004 Jan 3;**328**(7430):22–25.
9. Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. *J Natl Cancer Inst*. 1959;**22**:719–748.
10. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials*. 1986 Sep;**7**(3):177–188.
11. Ades AE. A chain of evidence with mixed comparisons: models for multiparameter synthesis and consistency of evidence. *Stat Med*. 2003 Oct 15;**22**(19):2995–3016.
12. Lu G, Ades AE. Combination of direct and indirect evidence in mixed treatment comparisons. *Stat Med*. 2004 Oct 30;**23**(20):3105–3124.
13. Caldwell DM, Ades AE, Higgins JPT. Simultaneous comparison of multiple treatments: combining direct and indirect evidence. *BMJ*. 2005 Oct 15;**331**(7521):897–900.
14. Spiegelhalter DJ, Best NG, Carlin BP, Van Der Linde, A. Bayesian measures of model complexity and fit. *J Roy Statist Soc B*. 2002;**64**(3):583–639.
15. Dempster AP. The direct use of likelihood for significance testing. *Stat Comp*. 1997;**7**:247–252.
16. Hamilton M. A rating scale for depression. *J Neurol Neurosurgery Psychiatry*. 1960;**23**:56–62.
17. Montgomery SA, Åsberg M. A new depression scale designed to be sensitive to change. *Br J Psychiatry*. 1979 Apr;**134**:382–389.
18. Philips Z, Ginnelly L, Sculpher M, Claxton K, Golder S, Riemsma R, *et al*. Review of guidelines for good practice in decision-analytic modelling in health technology assessment. *Health Technol Assess*. 2004 Sep;**8**(36):iii-iv, ix-xi, 1–158.
19. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ*. 2003 Sep 6;**327**(7414):557–560.
20. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ*. 1997 Sep 13;**315**(7109):629–634.
21. Evers S, Goossens M, de Vet H, van Tulder M, Ament A. Criteria list for assessment of methodological quality of economic evaluations: Consensus on Health Economic Criteria. *Int J Technol Assess Health Care*. 2005 Spring;**21**(2):240–245.
22. National Institute for Health and Clinical Excellence. *Guide to the methods of technology appraisal June 2008 (reference N1618)*. <http://www.nice.org.uk/media/B52/A7/TAMethodsGuideUpdatedJune2008.pdf>
23. *Age adjusted utility/adjusting utility for age Health Survey of England 1996*. <http://www.archive.official-documents.co.uk/document/doh/survey96/tab5-29.htm>

Appendix 2 Literature search strategies for the clinical review

Database: Ovid MEDLINE(R) In-Process & Other Non-Indexed Citations and Ovid MEDLINE(R) – 1946 to present

Search strategy

1. Depression/ or Depressive Disorder/ or Depressive Disorder, Major/ (126,194)
2. depress\$.tw. (267,648)
3. 1 or 2 (300,405)
4. Serotonin Uptake Inhibitors/ or Antidepressive Agents, Second-Generation/ or ssri\$.tw. or (serotonin adj2 inhibitor\$.tw. or Serotonin Antagonists/ (35,140)
5. (citalopram or citalopramum or celexa or cipramil).tw. or Citalopram/ (4259)
6. (escitalopram or S-citalopram or lexapro or ciprallex).tw. (939)
7. (fluoxetin\$ or fluokset\$ or felicium or prozac or prozep or prozit or sarafem or selfemra or symbyax).tw. or Fluoxetine/ (9738)
8. (fluvoxamin\$ or fluvoksami\$ or faverin or luvox).tw. or Fluvoxamine/ (2372)
9. (paroxetin\$ or parokset\$ or seroxat or paxil or pexeva).tw. or Paroxetine/ (4698)
10. (sertralin\$ or lustral or zoloft).tw. or Sertraline/ (3149)
11. 4 or 5 or 6 or 7 or 8 or 9 or 10 (42,763)
12. (refract\$ or resistan\$ or nonrespon\$ or unrespon\$ or fail\$ or (incomplet\$ adj respon\$) or (no\$ adj2 respon\$)).tw. (1,393,254)
13. treatment failure/ (22,171)
14. (inadequat\$ respon\$ or (sub\$ adj2 respon\$) or (poor\$ adj respon\$)).tw. (36,866)
15. 12 or 13 or 14 (1,427,769)
16. randomized controlled trial.pt. (314,314)
17. (random\$ or rct or placebo\$.tw. (623,289)
18. ((singl\$ or doubl\$ or tripl\$ or treb\$) adj (blind\$3 or mask\$3 or sham\$ or dummy)).tw. (112,536)
19. (case reports or comment or editorial or in vitro or letter).pt. (2,881,916)
20. 19 not (randomized controlled trial or controlled clinical trial).pt. (2,874,149)
21. (16 or 17 or 18) not 20 (693,670)
22. exp animals/ not humans.sh. (3,650,161)
23. 21 not 22 (621,736)
24. (anti-depress\$ or antidepress\$.tw. (41,423)
25. Antidepressive Agents/ (28,927)
26. 24 or 25 (54,177)
27. (augment\$ or adjunct\$ or combin\$ or add\$ or potentiation).tw. (2,822,627)
28. Drug Therapy, Combination/ (125,552)
29. 27 or 28 (2,888,000)
30. Lithium/ or exp Lithium Compounds/ (23,585)
31. antipsychotic agents/ or clozapine/ or risperidone/ (38,767)
32. (Lithium or Camcolit or Liskonum or Priadel or Li-Liquid or Litarex or Lithonate or Phasal or Lithny or Licio or Lithii or Litio or Litium or Lityum or dilithium or litu or Cibalith-S or Eskalith or Lithane or Lithobid or Lithonate or Lithotabs).tw. (25,357)
33. ((atypical adj antipsychotic?) or (atypical adj anti-psychotic?) or Amisulprid\$ or Solian or Aripiprazol\$ or Abilify or Clozapin\$ or Clozaril or Klotsapiini or Klozapin\$ or FazaClo or Fazalco or Denzapine or Zaponex or Olanzapin\$ or Olantsapiini or Zyprexa or Zypadhera or Paliperidon\$ or Invega or Xeplion or Quetiapin\$ or Seroquel or Risperidon\$ or Riszperidon or Rysperidon or Risperdal or Ziprasidon\$ or Geodon).tw. (19,696)

34. or/30-33 (74,184)
35. (26 and 29 and 34) or 11 (44,571)
36. 3 and 15 and 23 and 35 (938)

Database: PsycINFO – 1806 to August week 2 2011

Search strategy

1. exp Major Depression/ (75,749)
2. depress\$.tw. (180,819)
3. 1 or 2 (182,464)
4. Serotonin Antagonists/ or serotonin reuptake inhibitors/ or (ssri\$ or (serotonin adj2 inhibitor\$)).tw. (9804)
5. (citalopram or citalopramum or celexa or cipramil).tw. or Citalopram/ (1781)
6. (escitalopram or S-citalopram or lexapro or cipralext).tw. (615)
7. (fluoxetin\$ or fluokset\$ or feliciu\$ or prozac or prozep or prozit or sarafem or selfemra or symbyax).tw. or Fluoxetine/ (5066)
8. (fluvoxamin\$ or fluvoksami\$ or faverin or luvox).tw. or Fluvoxamine/ (1376)
9. (paroxetin\$ or parokset\$ or seroxat or paxil or pexeva).tw. or Paroxetine/ (2593)
10. (sertralini\$ or lustral or zoloft).tw. or Sertraline/ (1930)
11. 4 or 5 or 6 or 7 or 8 or 9 or 10 (16,639)
12. (refract\$ or resistan\$ or nonrespon\$ or unrespon\$ or fail\$ or (incomplet\$ adj respon\$) or (no\$ adj2 respon\$)).tw. (167,518)
13. (inadequat\$ respon\$ or (sub\$ adj2 respon\$) or (poor\$ adj respon\$)).tw. (8907)
14. 12 or 13 (174,891)
15. treatment outcome clinical trial.md. (19,559)
16. treatment effectiveness evaluation/ (12,499)
17. (random\$ or rct or placebo\$).tw. (117,674)
18. ((singl\$ or doubl\$ or tripl\$ or treb\$) adj (blind\$3 or mask\$3 or sham\$ or dummy)).tw. (16,252)
19. 15 or 16 or 17 or 18 (136,659)
20. (comment or editorial or letter).dt. (31,151)
21. 20 not treatment outcome clinical trial.md. (30,861)
22. 19 not 21 (135,679)
23. antidepressant drugs/ (13,273)
24. (antidepress\$ or anti-depress\$).tw. (26,031)
25. 23 or 24 (28,058)
26. (augment\$ or adjunct\$ or combin\$ or add\$ or potentiation).tw. (527,381)
27. polypharmacy/ (536)
28. Drug Augmentation/ (740)
29. 26 or 27 or 28 (527,592)
30. (Lithium or Camcolit or Liskonum or Priadel or Li-Liquid or Litarex or Lithonate or Phasal or Lithny or Licio or Lithii or Lito or Litium or Lityum or dilithium or litu or Cibalith-S or Eskalith or Lithane or Lithobid or Lithonate or Lithotabs).tw. (8478)
31. ((atypical adj antipsychotic?) or (atypical adj anti-psychotic?) or Amisulprid\$ or Solian or Aripiprazol\$ or Abilify or Clozapin\$ or Clozaril or Klotsapiini or Klozapin\$ or FazaClo or Fazalco or Denzapine or Zaponex or Olanzapin\$ or Olantsapiini or Zyprexa or Zypadhera or Paliperidon\$ or Invega or Xeplion or Quetiapin\$ or Seroquel or Risperidon\$ or Riszperidon or Rysperidon or Risperdal or Ziprasidon\$ or Geodon).tw. (14,792)
32. exp Lithium Carbonate/ or Lithium/ (5297)
33. neuroleptic drugs/ or aripiprazole/ or clozapine/ or olanzapine/ or quetiapine/ or risperidone/ (20,778)

34. 30 or 31 or 32 or 33 (30,510)
35. 25 and 29 and 34 (1550)
36. 11 or 35 (17,771)
37. 3 and 14 and 22 and 36 (720)

Database: EMBASE – 1974 to 2011 week 32

Search strategy

1. *depression/ or *major depression/ (107,356)
2. depress\$.tw. (302,295)
3. 1 or 2 (321,132)
4. serotonin antagonist/ or serotonin uptake inhibitor/ or fluoxetine plus olanzapine/ or fluvoxamine maleate/ or ssri\$.tw. or (serotonin adj2 inhibitor\$).tw. (46,279)
5. (citalopram or citalopramum or celexa or cipramil).tw. or citalopram/ (14,277)
6. (escitalopram or S-citalopram or lexapro or cipralext).tw. or escitalopram/ (4407)
7. (fluoxetin\$ or fluokset\$ or felicium or prozac or prozep or prozitet or sarafem or selfemra or symbyax).tw. or fluoxetine/ or fluoxetine plus olanzapine/ (32,341)
8. (fluvoxamin\$ or fluvoksami\$ or faverin or luvox).tw. or fluvoxamine/ or fluvoxamine maleate/ (10,747)
9. (paroxetin\$ or parokset\$ or seroxat or paxil or pexeva).tw. or paroxetine/ (19,872)
10. (sertralin\$ or lustral or zoloft).tw. or sertraline/ (16,373)
11. 4 or 5 or 6 or 7 or 8 or 9 or 10 (79,964)
12. (refract\$ or resistan\$ or nonrespon\$ or unrespon\$ or fail\$ or (incomplet\$ adj respon\$) or (no\$ adj2 respon\$)).tw. (1,520,915)
13. drug treatment failure/ (10,532)
14. (inadequat\$ respon\$ or (sub\$ adj2 respon\$) or (poor\$ adj respon\$)).tw. (40,713)
15. 12 or 13 or 14 (1,555,120)
16. randomized controlled trial/ (284,436)
17. (random\$ or rct or placebo\$).tw. (707,279)
18. ((singl\$ or doubl\$ or tripl\$ or treb\$) adj (blind\$3 or mask\$3 or sham\$ or dummy)).tw. (129,146)
19. 16 or 17 or 18 (787,496)
20. (editorial or letter).pt. (1,110,740)
21. 20 not (16 or clinical trial/) (1,070,423)
22. 19 not 21 (785,365)
23. exp animals/ not exp humans/ (1,237,993)
24. 22 not 23 (761,970)
25. (anti-depress\$ or antidepress\$).tw. (52,230)
26. antidepressant agent/ (59,124)
27. 25 or 26 (85,174)
28. (augment\$ or adjunct\$ or combin\$ or add\$ or potentiation).tw. (3,077,561)
29. drug combination/ (53,343)
30. 28 or 29 (3,112,487)
31. exp lithium/ or exp lithium derivative/ or lithium carbonate/ or lithium chloride/ or lithium citrate/ (46,132)
32. atypical antipsychotic agent/ or ziprasidone/ or risperidone/ or quetiapine/ or paliperidone/ or olanzapine/ or clozapine/ or clozapine derivative/ or clozapine n oxide/ or aripiprazole/ or amisulpride/ (46,689)
33. (Lithium or Camcolit or Liskonum or Priadel or Li-Liquid or Litarex or Lithonate or Phasal or Lithny or Licio or Lithii or Litio or Litium or Lityum or dilithium or litu or Cibalith-S or Eskalith or Lithane or Lithobid or Lithonate or Lithotabs).tw. (29,256)

34. ((atypical adj antipsychotic?) or (atypical adj anti-psychotic?) or Amisulprid\$ or Solian or Aripiprazol\$ or Abilify or Clozapin\$ or Clozaril or Klotsapiini or Klozapin\$ or FazaClo or Fazalco or Denzapine or Zaponex or Olanzapin\$ or Olantsapiini or Zyprexa or Zypadhera or Paliperidon\$ or Invega or Xeplion or Quetiapin\$ or Seroquel or Risperidon\$ or Riszperidon or Rysperydol or Risperdal or Ziprasidon\$ or Geodon).tw. (27,771)
35. or/31-34 (96,393)
36. (27 and 30 and 35) or 11 (81,812)
37. 3 and 15 and 24 and 36 (1303)

Database: Cochrane Central Register of Controlled Trials

Search strategy

- #1 MeSH descriptor Depressive Disorder, this term only (4140)
- #2 MeSH descriptor Depression, this term only (4118)
- #3 depress*:ti,ab in Clinical Trials (25,500)
- #4 MeSH descriptor Depressive Disorder, Major, this term only (1781)
- #5 MeSH descriptor Antidepressive Agents, Second-Generation, this term only (1065)
- #6 MeSH descriptor Serotonin Antagonists, this term only (878)
- #7 MeSH descriptor Serotonin Uptake Inhibitors, this term only (2018)
- #8 MeSH descriptor Citalopram, this term only (606)
- #9 MeSH descriptor Fluoxetine, this term only (1086)
- #10 MeSH descriptor Fluvoxamine, this term only (347)
- #11 MeSH descriptor Paroxetine, this term only (697)
- #12 MeSH descriptor Sertraline, this term only (547)
- #13 ssri* or (serotonin near/2 inhibitor*) or citalopram or citalopramum or celexa or cipramil or escitalopram or S-citalopram or lexapro or cipralext or fluoxetin* or fluokset* or feliciu or prozac or prozep or prozit or sarafem or selfemra or symbyax or fluvoxamin* or fluvoksami* or faverin or luvox or paroxetin* or parokset* or seroxat or paxil or pexeva or sertralin* or lustral or zoloft in Clinical Trials (6667)
- #14 (#5 OR #6 OR #7 OR #8 OR #9 OR #10 OR #11 OR #12 OR #13) (7660)
- #15 refract* or resistan* or nonrespon* or (no* near/2 respon*) or unrespons* or fail* or (incomplet* near/1 respon*) or (inadequat* near/1 respons*) or (sub* near/2 respon*) or (poor* near/1 respon*): ti,ab in Clinical Trials (75,271)
- #16 MeSH descriptor Treatment Failure, this term only (2311)
- #17 (#15 OR #16) (75,271)
- #18 antidepress* or anti-depress*:ti,ab in Clinical Trials (7707)
- #19 MeSH descriptor Antidepressive Agents, this term only (2441)
- #20 augment* or adjunct* or combin* or add* or potentiation:ti,ab in Clinical Trials (365,207)
- #21 MeSH descriptor Drug Therapy, Combination, this term only (22,330)

- #22 Lithium or Camcolit or Liskonum or Priadel or Li-Liquid or Litarex or Lithonate or Phasal or Lithny or Licio or Lithii or Lio or Litium or Lityum or dilithium or litu or Cibalith-S or Eskalith or Lithane or Lithobid or Lithonate or Lithotabs in Clinical Trials (1579)
- #23 (atypical near/1 antipsychotic*) or (atypical near/1 anti-psychotic*) or Amisulprid* or Solian or Aripiprazol* or Abilify or Clozapin* or Clozaril or Klotsapiini or Klozapin* or FazaClo or Fazalco or Denzapine or Zaponex or Olanzapin* or Olantsapiini or Zyprexa or Zypadhera or Paliperidon* or Invega or Xeplion or Quetiapin* or Seroquel or Risperidon* or Riszperidon or Rysperrydon or Risperdal or Ziprasidon* or Geodon in Clinical Trials (4083)
- #24 MeSH descriptor Lithium, this term only (626)
- #25 MeSH descriptor Lithium Compounds explode all trees (349)
- #26 MeSH descriptor Lithium Carbonate, this term only (239)
- #27 MeSH descriptor Lithium Chloride, this term only (20)
- #28 MeSH descriptor Antipsychotic Agents, this term only (3068)
- #29 MeSH descriptor Clozapine, this term only (378)
- #30 MeSH descriptor Risperidone, this term only (733)
- #31 (#22 OR #23 OR #24 OR #25 OR #26 OR #27 OR #28 OR #29 OR #30) (6535)
- #32 ((#18 OR #19) AND (#20 OR #21) AND #31) (403)
- #33 ((#1 OR #2 OR #3 OR #4) AND (#14 OR #32) AND #17) (756)

Appendix 3 Copy of data extraction form

TREATMENT-RESISTANT DEPRESSION DATA EXTRACTION FORM

PART ONE: REVIEW, REVIEWER AND STUDY INFORMATION

Study ID:

Reviewer name:

Date of completion of this form:

Title of paper/abstract:

Source (journal, year, volume, pages):

Authors:

Language of publication:

Type of paper (e.g. full paper/abstract/poster):

PART TWO: VERIFICATION OF STUDY ELIGIBILITY

Type of clinical trial

1) Is the study randomised? YES UNCLEAR NO

Population in the clinical trial

2) Is the population adults ≥ 18 years old? YES UNCLEAR NO

3) Did the RCT include people with unipolar depression? YES UNCLEAR NO

4) Did the RCT include people with treatment resistant depression (defined as failure to respond to ≥ 2 antidepressants)? YES UNCLEAR NO

Interventions in the clinical trial

5) Does the trial compare SSRI + atypical antipsychotic or lithium or no treatment with SSRI + lithium or atypical antipsychotic or placebo or no treatment? YES UNCLEAR NO

6) Did both groups experience the same care except for the two interventions under investigation? YES UNCLEAR NO

Outcomes of the clinical trial

7) Does the study report on outcomes during the treatment of the acute phase of depression? YES UNCLEAR NO

8) Did the clinical trial investigate at least one of the following: disease severity, quality of life, adverse events, withdrawals (all cause), relapse rate, mortality (all cause)? YES UNCLEAR NO

9) Are the outcomes measured after ≥ 4 weeks treatment with study medication? YES UNCLEAR NO

If you answered NO to any of the above questions do not proceed to Part 3.

PART THREE: INFORMATION ABOUT THE STUDY**Characteristics of the trial**

Country(ies) where the clinical trial was conducted:

Sponsors of the clinical trial:

Any conflicts of interest reported for any of the researchers?

Date the clinical trial was conducted:

Type of clinical trial design (e.g. parallel, crossover, or cluster trial):

If the trial was of crossover design, are there pre-crossover results reported?

Was the trial multicentre? If so, how many centres were there?

Characteristics of the patients*Inclusion criteria*

- How and where were patients enrolled, were any patient risk factors used?
- What details of the antidepressant(s) patients had failed to respond to are provided?
- How was treatment resistance defined?

Exclusion criteria

- Were specific groups of people excluded?

Total number of people randomised:

Information on the age of the patients:

Information on the sex of the patients (m/f):

Information on the ethnicity of the patients:

Information on patients' medical history (i.e. previous depression):

Type of intervention:**Intervention 1 SSRI + XX (where XX = atypical antipsychotic or lithium or no treatment)**

SSRI name and brand:

SSRI dose and regimen used [e.g. 80 mg once daily (OD)]:

Delivery of SSRI [e.g. per os (p.o.) tablet/dissolvable/enteric coated]:

Number of doses of SSRI given per day (with SD/SE if given):

Duration of SSRI treatment in days (with SD/SE if given):

What was XX (name and brand)?

XX dose and regimen used (e.g. 80 mg OD):

Delivery of XX (e.g. p.o. tablet/dissolvable/enteric coated):

Number of doses of XX given per day (with SD/SE if given):

Duration of XX treatment in days (with SD/SE if given):

Number of patients randomised:

Intervention 2 SSRI + YY (where YY = lithium or atypical antipsychotic or placebo or no treatment)

SSRI name and brand:

SSRI dose and regimen used (e.g. 80 mg OD):

Delivery of SSRI (e.g. p.o. tablet/dissolvable/enteric coated):

Number of doses of SSRI given per day (with SD/SE if given):

Duration of SSRI treatment in days (with SD/SE if given):

What was YY (name and brand)?

YY dose and regimen used (e.g. 80 mg OD):

Delivery of YY (e.g. p.o. tablet/dissolvable/enteric coated):

Number of doses of YY given per day (with SD/SE if given):

Duration of YY treatment in days (with SD/SE if given):

Number of patients randomised:

Was the formulation and appearance of YY (e.g. lithium) matched to that of XX (e.g. atypical antipsychotic)?

Were any additional interventions given to either or both groups?

Types of outcome

Which of the following outcomes have been assessed in the clinical trial?

Response?

How was response defined in the clinical trial? YES UNCLEAR NO

QoL?

How was QoL defined in the clinical trial? YES UNCLEAR NO

Adverse events?

How were adverse events defined in the clinical trial? (e.g. investigator attributed?) YES UNCLEAR NO

Withdrawal (all cause)?

How was withdrawal defined in the clinical trial? YES UNCLEAR NO

Relapse rate?

How was relapse rate defined in the clinical trial? YES UNCLEAR NO

All-cause mortality?

How was all-cause mortality defined in the clinical trial? YES UNCLEAR NO

Remission rate?

How was remission rate defined in the clinical trial? YES UNCLEAR NO

Other outcomes

Any other outcomes reported in trial (please list)? YES UNCLEAR NO

ITT data collection table

Outcomes		Time frame (weeks)	SSRI + olanzapine		SSRI + placebo	
			<i>n</i>	<i>N</i>	<i>n</i>	<i>N</i>
Response	≥ 50% reduction in MADRS					
	≥ 50% reduction in HAMD					
	Improvement from baseline on MADRS					
	Improvement from baseline on CGI					
	Improvement from baseline on HAMA					
	Improvement from baseline on BPRS					
QoL	Trial scale					
Withdrawals (all cause)						
Relapse rate						
Remission rate						
All-cause mortality						
Adverse events (please specify)	Weight gain					

BPRS, Brief Psychiatric Rating Scale; HAMA, Hamilton Anxiety Scale; *n*, number of patients with the outcome; *N*, number of patients assessed.

Per-protocol data collection table

Outcomes		Time frame (weeks)	SSRI + olanzapine		SSRI + placebo	
			<i>n</i>	<i>N</i>	<i>n</i>	<i>N</i>
Response	≥ 50% reduction in MADRS					
	≥ 50% reduction in HAMD					
	Improvement from baseline on MADRS					
	Improvement from baseline on CGI					
	Improvement from baseline on HAMA					
	Improvement from baseline on BPRS					
QoL	Trial scale					
Withdrawals (all cause)						
Relapse rate						
Remission rate						
All-cause mortality						
Adverse events (please specify)	Weight gain					

BPRS, Brief Psychiatric Rating Scale; HAMA, Hamilton Anxiety Scale; *n*, number of patients with the outcome; *N*, number of patients assessed.

Did the RCT carry out any subgroup analyses of interest? (i.e. Different durations of depression, different classes of previous antidepressants, different genders, age, different severities of depression or different number of prior antidepressants.)

If yes, please give details here:

Did the RCT provide any details of maintenance therapy +/- outcomes?

PART FOUR: CLINICAL TRIAL QUALITY

Please describe the method of randomisation and allocation concealment used in the clinical trial:

Please describe the method of blinding and who was blinded in the clinical trial:

How would you describe the trial's design to minimise bias for (please tick):

Outcome	Risk of bias	Low risk	Unclear risk	High risk	Comments
Response	1) Random sequence generation				
	2) Allocation concealment				
	3) Blinding (participants and personnel)				
	4) Blinding of outcomes assessment				
	5) Incomplete outcome data				
	6) Selective reporting				
	7) 'Other bias'				
QoL	1) Random sequence generation				
	2) Allocation concealment				
	3) Blinding (participants and personnel)				
	4) Blinding of outcomes assessment				
	5) Incomplete outcome data				
	6) Selective reporting				
	7) 'Other bias'				
Withdrawals (all cause)	1) Random sequence generation				
	2) Allocation concealment				
	3) Blinding (participants and personnel)				
	4) Blinding of outcomes assessment				
	5) Incomplete outcome data				
	6) Selective reporting				
	7) 'Other bias'				
Relapse	1) Random sequence generation				
	2) Allocation concealment				
	3) Blinding (participants and personnel)				
	4) Blinding of outcomes assessment				
	5) Incomplete outcome data				
	6) Selective reporting				
	7) 'Other bias'				
Remission	1) Random sequence generation				
	2) Allocation concealment				
	3) Blinding (participants and personnel)				
	4) Blinding of outcomes assessment				
	5) Incomplete outcome data				
	6) Selective reporting				
	7) 'Other bias'				

Outcome	Risk of bias	Low risk	Unclear risk	High risk	Comments
All-cause mortality	1) Random sequence generation				
	2) Allocation concealment				
	3) Blinding (participants and personnel)				
	4) Blinding of outcomes assessment				
	5) Incomplete outcome data				
	6) Selective reporting				
	7) 'Other bias'				
Adverse events	1) Random sequence generation				
	2) Allocation concealment				
	3) Blinding (participants and personnel)				
	4) Blinding of outcomes assessment				
	5) Incomplete outcome data				
	6) Selective reporting				
	7) 'Other bias'				

How would you rate the trial's overall risk of bias?

Low risk Unclear High risk

Do you have any additional comments you would like to make about this clinical trial?

Ideally, would you like further information about the clinical trial from the authors (if so, please give details)?

Appendix 4 Table of excluded clinical effectiveness studies with rationale

Reference details	Reason for exclusion
Austin MP, Souza FG, Goodwin GM. Lithium augmentation in antidepressant-resistant patients. A quantitative analysis. <i>Br J Psychiatry</i> 1991; 159 :510–14	Systematic review
Bauer M, Dopfmer S. Lithium augmentation in treatment-resistant depression: meta-analysis of placebo-controlled studies. <i>J Clin Psychopharmacol</i> 1999; 19 :427–34	Systematic review
Bauer M, Adli M, Baethge C, Berghofer A, Sasse J, Heinz A, et al. Lithium augmentation therapy in refractory depression: clinical evidence and neurobiological mechanisms. <i>Can J Psychiatr Rev Canad Psychiatr</i> 2003; 48 :440–8	Systematic review
Bauer M, Forsthoef A, Baethge C, Adli M, Berghofer A, Dopfmer S, et al. Lithium augmentation therapy in refractory depression-update 2002. <i>Eur Arch Psychiatr Clin Neurosci</i> 2003; 253 :132–9	Systematic review
Bauer M, Pretorius HW, Constant EL, Earley WR, Szamosi J, Brecher M, et al. Extended-release quetiapine as adjunct to an antidepressant in patients with major depressive disorder: results of a randomized, placebo-controlled, double-blind study. <i>J Clin Psychiatr</i> 2009; 70 :540–9	Did not meet population or intervention inclusion criteria
Bauer M, Adli M, Bschor T, Pilhatsch M, Pfennig A, Sasse J, et al. Lithium's emerging role in the treatment of refractory major depressive episodes: augmentation of antidepressants. <i>Neuropsychobiology</i> 2010; 62 :36–42	Systematic review
Baumann P, Nil R, Souche A, Montaldi S, Baettig D, Lambert S, et al. A double-blind, placebo-controlled study of citalopram with and without lithium in the treatment of therapy-resistant depressive patients: a clinical, pharmacokinetic, and pharmacogenetic investigation. <i>J Clin Psychopharmacol</i> 1996; 16 :307–14	Did not meet population or treatment duration inclusion criteria
Bin W, Mei H, Po L. Fluoxetine and olanzapine in the treatment of refractory depression. <i>Shandong Arch Psychiatry</i> 2006; 19 :87–9	Unable to obtain a print copy
Bobo WV, Shelton RC. Fluoxetine and olanzapine combination therapy in treatment-resistant major depression: review of efficacy and safety data. <i>Exp Opin Pharmacother</i> 2009; 10 :2145–59	Systematic review
Boulton DW, Balch AH, Royzman K, Patel CG, Berman RM, Mallikaarjun S, et al. The pharmacokinetics of standard antidepressants with aripiprazole as adjunctive therapy: studies in healthy subjects and in patients with major depressive disorder. <i>J Psychopharmacol</i> 2010; 24 :537–46	Did not report any outcomes of interest
Browne M, Lapierre YD, Hrdina PD, Horn E. Lithium as an adjunct in the treatment of major depression. <i>Int Clin Psychopharmacol</i> 1990; 5 :103–10	Did not meet population or intervention inclusion criteria
Bschor T, Bauer M. Efficacy and mechanisms of action of lithium augmentation in refractory major depression. <i>Curr Pharm Des</i> 2006; 12 :2985–92	Systematic review
Cooper C, Katona C, Lyketsos K, Blazer D, Brodaty H, Rabins P, et al. A systematic review of treatments for refractory depression in older people. <i>Am J Psychiatr</i> 2011; 168 :681–8	Systematic review
Crossley NA, Bauer M. Acceleration and augmentation of antidepressants with lithium for depressive disorders: two meta-analyses of randomized, placebo-controlled trials. <i>J Clin Psychiatr</i> 2007; 68 :935–40	Systematic review

Reference details	Reason for exclusion
Denko TCF. Augmentation strategies in STAR*D: a review. <i>Prim Psychiatr</i> 2007; 14 :46–50	Systematic review
Dorée JP, Des Rosiers J, Lew V, Gendronc A, Elied R, Stipe E, <i>et al.</i> Quetiapine augmentation of treatment-resistant depression: a comparison with lithium. <i>Curr Med Res Opin</i> 2007; 23 :333–41	Did not meet population or intervention inclusion criteria
Dube S. Meta-analysis of olanzapine-fluoxetine in treatment-resistant depression. 155th Annual Meeting of the American Psychiatric Association; 18–23 May 2002, Philadelphia, PA, USA	Unable to obtain and a systematic review
Dube S, Dube S, Andersen SW, Corya SA, Sanger TM, Tollefson GD. Olanzapine-fluoxetine for treatment-resistant depression. XII World Congress of Psychiatry, 24–29 August 2002, Yokohama, Japan	Unable to obtain a print copy
Dunner DL, Amsterdam JD, Shelton RC, Loebel A, Romano SJ. Efficacy and tolerability of adjunctive ziprasidone in treatment-resistant depression: a randomized, open-label, pilot study. <i>J Clin Psychiatr</i> 2007; 68 :1071–7	Decision to exclude based on high doses of ziprasidone and not used in the UK
El-Khalili N, Joyce M, Atkinson S, Buynak RJ, Datto C, Lindgren P, <i>et al.</i> Extended-release quetiapine fumarate (quetiapine XR) as adjunctive therapy in major depressive disorder (MDD) in patients with an inadequate response to ongoing antidepressant treatment: a multicentre, randomized, double-blind, placebo-controlled study. <i>Int J Neuropsychopharmacol</i> 2010; 13 :917–32	Did not meet population or intervention inclusion criteria
Fava M, Rosenbaum JF, McGrath PJ, Stewart JW, Amsterdam JD, Quitkin FM. Lithium and tricyclic augmentation of fluoxetine treatment for resistant major depression: a double-blind, controlled study. <i>Am J Psychiatr</i> 1994; 151 :1372–4	Did not meet population or intervention inclusion criteria
Fava M, Alpert J, Nierenberg A, Lagomasino I, Sonawalla S, Tedlow J, <i>et al.</i> Double-blind study of high-dose fluoxetine vs lithium or desipramine augmentation of fluoxetine in partial responders and nonresponders to fluoxetine. <i>J Clin Psychopharmacol</i> 2002; 22 :379–87	Did not meet population or intervention inclusion criteria
Fleurence R, Williamson R, Jing Y, Kim E, Tran QV, Pikalov AS, <i>et al.</i> A systematic review of augmentation strategies for patients with major depressive disorder. <i>Psychopharmacol Bull</i> 2009; 42 :57–90	Unable to obtain a print copy
Kanto D, McNeven S, Leichner P, Harper D, Krenn M. The benefit of lithium carbonate adjunct in refractory depression – fact or fiction? <i>Can J Psychiatr Rev Canad Psychiatr</i> 1986; 31 :416–18	Did not meet population or intervention inclusion criteria
Keitner GI, Garlow SJ, Ryan CE, Ninan PT, Solomon DA, Nemeroff CB, <i>et al.</i> A randomized, placebo-controlled trial of risperidone augmentation for patients with difficult-to-treat unipolar, non-psychotic major depression. <i>J Psychiatr Res</i> 2009; 43 :205–14	Did not meet population or intervention inclusion criteria
Kennedy SH, Lam RW, Cohen NL, Ravindran AV. Clinical guidelines for the treatment of depressive disorders. IV. Medications and other biological treatments. <i>Can J Psychiatr Rev Canad Psychiatr</i> 2001; 46 (Suppl. 1):38–58	Systematic review
Li H, Zhang Y, Zhang Y. Double-blind study of fluoxetine augmented with olanzapine in the treatment of treatment-resistant depression. <i>Shandong Arch Psychiatr</i> 2006; 19 :85–6	Unable to obtain a print copy
Mahmoud RA, Pandina GJ, Turkoz I, Kosik-Gonzalez C, Canuso CM, Kujawa MJ, <i>et al.</i> Risperidone for treatment-refractory major depressive disorder: a randomized trial. <i>Ann Int Med</i> 2007; 147 :593–602	Did not meet population or intervention inclusion criteria
Nelson JC, Papakostas GI. Atypical antipsychotic augmentation in major depressive disorder: a meta-analysis of placebo-controlled randomized trials. <i>Am J Psychiatr</i> 2009; 166 :980–91	Systematic review
Papakostas GI, Shelton RC, Smith J, Fava M. Augmentation of antidepressants with atypical antipsychotic medications for treatment-resistant major depressive disorder: a meta-analysis. <i>J Clin Psychiatr</i> 2007; 68 :826–31	Systematic review

Reference details	Reason for exclusion
Patten SB, Lupin DA, Boucher SA, Lamarre CJ. Pharmacologic management of refractory depression. <i>CMAJ</i> 1992; 146 :483–7	Systematic review
Philip NS, Carpenter LL, Tyrka AR, Price LH. Augmentation of antidepressants with atypical antipsychotics: a review of the current literature. <i>J Psychiatr Pract</i> 2008; 14 :34–44	Systematic review
Price LHC. Lithium augmentation for refractory depression: a critical reappraisal. <i>Prim Psychiatr</i> 2008; 15 :35–42	Systematic review
Rapaport MH, Gharabawi GM, Canuso CM, Mahmoud RA, Keller MB, Bossie CA, <i>et al.</i> Effects of risperidone augmentation in patients with treatment-resistant depression: Results of open-label treatment followed by double-blind continuation. <i>Neuropsychopharmacology</i> 2006; 31 :2505–13	Did not meet inclusion criteria of acute phase treatment
Rouillon F, Gorwood P. The use of lithium to augment antidepressant medication. <i>J Clin Psychiatr</i> 1998; 59 :32–41	Systematic review
Santos MAH. Treatment-resistant depression: review of pharmacologic antidepressant strategies. <i>J Bras Psiquiatr</i> 2006; 55 :232–42	Systematic review
Savas HAK. Aripiprazole in the treatment of depression: a review. <i>Klinik Psikofarmakoloji Bulteni</i> 2010; 20 :S26–30	Systematic review
Schmauss M, Erfurth A. [Combination therapies in antidepressive drug refractory depression: an overview.] <i>Fortschr Neurologie Psych</i> 1996; 64 :390–402	Systematic review
Schmauss M, Messer T. [Augmentation strategies for therapy resistant depression – a review.] <i>Psychiatr Prax</i> 2007; 34 :165–74	Systematic review
Schweitzer I, Tuckwell V, Johnson G. A review of the use of augmentation therapy for the treatment of resistant depression: implications for the clinician. <i>Aust NZ J Psychiatr</i> 1997; 31 :340–52	Systematic review
Selis MA, Peeters FP. [Augmentation with atypical antipsychotics for the treatment of patients with a therapy-resistant depression: a review.] <i>Tijdschr Psychiatr</i> 2008; 50 :213–22	Systematic review
Souche A, Montaldi S, Uehlinger C, Kasas A, Reymond MJ, Reymond P, <i>et al.</i> [Treatment of resistant depression with the citalopram–lithium combination. Methodology of a double-blind multicenter study and preliminary results.] <i>Encephale</i> 1991; 17 :213–19	Did not meet treatment duration inclusion criteria
Stefanescu C, Mavros M, Chirita V. Resistant depression: a comparison between antidepressants and the efficacy of olanzapine augmentation. <i>Psychiatriki</i> 2005; 16 :252	Unable to obtain a print copy
Stimpson N, Agrawal N, Lewis G. Randomised controlled trials investigating pharmacological and psychological interventions for treatment-refractory depression. Systematic review. <i>Br J Psychiatry</i> 2002; 181 :284–94	Systematic review
Tollefson G, Gannon K, Jacobs T, Shelton R, Tohen M, Stahl S. The study of olanzapine plus fluoxetine in treatment-resistant major depressive disorder without psychotic features. XI World Congress of Psychiatry, 6–11 August 1999, Hamburg, Germany, Abstracts Volume II:133	Unable to obtain a print copy
Trivedi MH, Thase ME, Osuntokun O, Henley DB, Case M, Watson SB, <i>et al.</i> An integrated analysis of olanzapine/fluoxetine combination in clinical trials of treatment-resistant depression. <i>J Clin Psychiatr</i> 2009; 70 :387–96	Systematic review
Tyrer P, Marsden CA, Casey P, Seivewright N. Clinical efficacy of paroxetine in resistant depression. <i>J Psychopharmacol</i> 1987; 1 :251–7	Unable to obtain a print copy

Reference details	Reason for exclusion
Wang X-H, Guo X. Effectiveness and safety of olanzapine combined with fluoxetine for refractory depression: a systematic review. <i>Chin J Evid-Based Med</i> 2010; 10 :1102–9	Systematic review
Zhornitsky S, Potvin S, Motesafi H, Dubreucq S, Rompre PP, Stip E. Dose-response and comparative efficacy and tolerability of quetiapine across psychiatric disorders: a systematic review of the placebo-controlled monotherapy and add-on trials. <i>Int Clin Psychopharmacol</i> 2011; 26 :183–92	Systematic review
Zusky PM, Biederman J, Rosenbaum JF, et al. Adjunct low dose lithium carbonate in treatment-resistant depression: a placebo-controlled study. <i>J Clin Psychopharmacol</i> 1988; 8 :120–4	Did not meet population or intervention inclusion criteria

Appendix 5 Quality assessment of clinical trials

Please note that the risk-of-bias analyses reported below are reported only for the outcomes for which the trial has been included in an analysis for that outcome.

Berman *et al.* 2007⁴⁷

Outcome	Risk of bias	Low risk	Unclear risk	High risk	Comments
Response	Random sequence generation		X		Limited details reported on trial methodology
	Allocation concealment		X		
	Blinding (participants and personnel)		X		
	Blinding of outcomes assessment		X		
	Incomplete outcome data		X		
	Selective reporting		X		
	'Other bias'		X		
Mean change in MADRS	Random sequence generation		X		Limited details reported on trial methodology
	Allocation concealment		X		
	Blinding (participants and personnel)		X		
	Blinding of outcomes assessment		X		
	Incomplete outcome data		X		
	Selective reporting		X		
	'Other bias'		X		
Withdrawals (all cause)	Random sequence generation		X		Limited details reported on trial methodology
	Allocation concealment		X		
	Blinding (participants and personnel)		X		
	Blinding of outcomes assessment		X		
	Incomplete outcome data		X		
	Selective reporting		X		
	'Other bias'		X		
Overall trial risk-of-bias rating:					
Low risk	Unclear			High risk	

Berman *et al.* 2009⁴⁸

Outcome	Risk of bias	Low risk	Unclear risk	High risk	Comments
Response	Random sequence generation		X		Limited details reported on trial methodology
	Allocation concealment		X		
	Blinding (participants and personnel)		X		
	Blinding of outcomes assessment		X		
	Incomplete outcome data		X		
	Selective reporting		X		
	'Other bias'		X		
Mean change in MADRS	Random sequence generation		X		Limited details reported on trial methodology
	Allocation concealment		X		
	Blinding (participants and personnel)		X		
	Blinding of outcomes assessment		X		
	Incomplete outcome data		X		
	Selective reporting		X		
	'Other bias'		X		
Withdrawals (all cause)	Random sequence generation		X		Limited details reported on trial methodology
	Allocation concealment		X		
	Blinding (participants and personnel)		X		
	Blinding of outcomes assessment		X		
	Incomplete outcome data		X		
	Selective reporting		X		
	'Other bias'		X		
Overall trial risk-of-bias rating:					
Low risk	Unclear			High risk	

Corya et al. 2006⁴⁹

Outcome	Risk of bias	Low risk	Unclear risk	High risk	Comments
Response	Random sequence generation		X		Limited details reported on trial methodology. Potential issue of non-blinding of investigators at randomisation: patients might not be as treatment resistant as in other trials
	Allocation concealment		X		
	Blinding (participants and personnel)		X		
	Blinding of outcomes assessment		X		
	Incomplete outcome data		X		
	Selective reporting		X		
	'Other bias'		X		
Mean change in MADRS	Random sequence generation		X		Limited details reported on trial methodology. Potential issue of non-blinding of investigators at randomisation: patients might not be as treatment resistant as in other trials
	Allocation concealment		X		
	Blinding (participants and personnel)		X		
	Blinding of outcomes assessment		X		
	Incomplete outcome data		X		
	Selective reporting		X		
	'Other bias'		X		
Remission	Random sequence generation		X		Limited details reported on trial methodology. Potential issue of non-blinding of investigators at randomisation: patients might not be as treatment resistant as in other trials
	Allocation concealment		X		
	Blinding (participants and personnel)		X		
	Blinding of outcomes assessment		X		
	Incomplete outcome data		X		
	Selective reporting		X		
	'Other bias'		X		
Withdrawals (all cause)	Random sequence generation		X		Limited details reported on trial methodology. Potential issue of non-blinding of investigators at randomisation: patients might not be as treatment resistant as in other trials
	Allocation concealment		X		
	Blinding (participants and personnel)		X		
	Blinding of outcomes assessment		X		
	Incomplete outcome data	X			
	Selective reporting		X		
	'Other bias'		X		
Adverse events	Random sequence generation		X		Limited details reported on trial methodology. Potential issue of non-blinding of investigators at randomisation: patients might not be as treatment resistant as in other trials
	Allocation concealment		X		
	Blinding (participants and personnel)		X		
	Blinding of outcomes assessment		X		
	Incomplete outcome data		X		
	Selective reporting		X		
	'Other bias'		X		
Overall trial risk-of-bias rating:					
Low risk	Unclear		High risk		

Feng et al. 2008⁴³

Outcome	Risk of bias	Low risk	Unclear risk	High risk	Comments
Response	Random sequence generation		X		Limited details reported on trial methodology
	Allocation concealment		X		
	Blinding (participants and personnel)		X		
	Blinding of outcomes assessment		X		
	Incomplete outcome data	X			
	Selective reporting		X		
	'Other bias'		X		
Adverse events	Random sequence generation		X		Limited details reported on trial methodology
	Allocation concealment		X		
	Blinding (participants and personnel)		X		
	Blinding of outcomes assessment		X		
	Incomplete outcome data		X		
	Selective reporting		X		
	'Other bias'		X		
Remission rate	Random sequence generation		X		Limited details reported on trial methodology
	Allocation concealment		X		
	Blinding (participants and personnel)		X		
	Blinding of outcomes assessment		X		
	Incomplete outcome data	X			
	Selective reporting		X		
	'Other bias'		X		
Overall trial risk-of-bias rating:					
Low risk	Unclear			High risk	

Franco *et al.* 2010³⁰

Outcome	Risk of bias	Low risk	Unclear risk	High risk	Comments
Response	Random sequence generation		X		Limited details reported on trial methodology
	Allocation concealment		X		
	Blinding (participants and personnel)		X		
	Blinding of outcomes assessment		X		
	Incomplete outcome data		X		
	Selective reporting		X		
	'Other bias'		X		
Mean change in MADRS	Random sequence generation		X		Limited details reported on trial methodology
	Allocation concealment		X		
	Blinding (participants and personnel)		X		
	Blinding of outcomes assessment		X		
	Incomplete outcome data		X		
	Selective reporting		X		
	'Other bias'		X		
Overall trial risk-of-bias rating:					
Low risk	Unclear			High risk	

Katona *et al.* 1995⁵⁹

Outcome	Risk of bias	Low risk	Unclear risk	High risk	Comments
Response	Random sequence generation		X		Limited details reported on trial methodology
	Allocation concealment		X		
	Blinding (participants and personnel)	X			
	Blinding of outcomes assessment		X		
	Incomplete outcome data		X		
	Selective reporting		X		
	'Other bias'		X		
Mean change in MADRS	Random sequence generation		X		Limited details reported on trial methodology
	Allocation concealment		X		
	Blinding (participants and personnel)	X			
	Blinding of outcomes assessment		X		
	Incomplete outcome data		X		
	Selective reporting		X		
	'Other bias'		X		

Outcome	Risk of bias	Low risk	Unclear risk	High risk	Comments
Withdrawals (all cause)	Random sequence generation		X		Limited details reported on trial methodology
	Allocation concealment		X		
	Blinding (participants and personnel)	X			
	Blinding of outcomes assessment		X		
	Incomplete outcome data		X		
	Selective reporting		X		
	'Other bias'		X		
Adverse events	Random sequence generation		X		Limited details reported on trial methodology
	Allocation concealment		X		
	Blinding (participants and personnel)	X			
	Blinding of outcomes assessment		X		
	Incomplete outcome data		X		
	Selective reporting		X		
	'Other bias'		X		
Overall trial risk-of-bias rating:					
Low risk	Unclear			High risk	

Mattingly *et al.* 2006⁴⁶

Outcome	Risk of bias	Low risk	Unclear risk	High risk	Comments
Response	Random sequence generation		X		Limited details reported on trial methodology
	Allocation concealment		X		
	Blinding (participants and personnel)		X		
	Blinding of outcomes assessment		X		
	Incomplete outcome data		X		
	Selective reporting		X		
	'Other bias'		X		
Mean change in MADRS	Random sequence generation		X		Limited details reported on trial methodology
	Allocation concealment		X		
	Blinding (participants and personnel)		X		
	Blinding of outcomes assessment		X		
	Incomplete outcome data		X		
	Selective reporting		X		
	'Other bias'		X		
Overall trial risk-of-bias rating:					
Low risk	Unclear			High risk	

Marcus *et al.* 2008⁵⁰

Outcome	Risk of bias	Low risk	Unclear risk	High risk	Comments
Response	Random sequence generation		X		Limited details reported on trial methodology
	Allocation concealment		X		
	Blinding (participants and personnel)		X		
	Blinding of outcomes assessment		X		
	Incomplete outcome data		X		
	Selective reporting		X		
	'Other bias'		X		
Mean change in MADRS	Random sequence generation		X		Limited details reported on trial methodology
	Allocation concealment		X		
	Blinding (participants and personnel)		X		
	Blinding of outcomes assessment		X		
	Incomplete outcome data		X		
	Selective reporting		X		
	'Other bias'		X		
Withdrawals (all cause)	Random sequence generation		X		Limited details reported on trial methodology
	Allocation concealment		X		
	Blinding (participants and personnel)		X		
	Blinding of outcomes assessment		X		
	Incomplete outcome data		X		
	Selective reporting		X		
	'Other bias'		X		
Overall trial risk-of-bias rating:					
Low risk	Unclear			High risk	

Shelton *et al.* 2001⁵¹

Outcome	Risk of bias	Low risk	Unclear risk	High risk	Comments
Response	Random sequence generation		X		Limited details reported on trial methodology
	Allocation concealment		X		
	Blinding (participants and personnel)		X		
	Blinding of outcomes assessment		X		
	Incomplete outcome data		X		
	Selective reporting		X		
	'Other bias'		X		

Outcome	Risk of bias	Low risk	Unclear risk	High risk	Comments
Withdrawals (all cause)	Random sequence generation		X		Limited details reported on trial methodology
	Allocation concealment		X		
	Blinding (participants and personnel)		X		
	Blinding of outcomes assessment		X		
	Incomplete outcome data		X		
	Selective reporting		X		
	'Other bias'		X		
Adverse events	Random sequence generation		X		Limited details reported on trial methodology
	Allocation concealment		X		
	Blinding (participants and personnel)		X		
	Blinding of outcomes assessment		X		
	Incomplete outcome data		X		
	Selective reporting		X		
	'Other bias'		X		
Overall trial risk-of-bias rating:					
Low risk	Unclear			High risk	

Shelton *et al.* 2005⁵²

Outcome	Risk of bias	Low risk	Unclear risk	High risk	Comments
Response	Random sequence generation		X		Limited details reported on trial methodology. Potential issue of non-blinding of investigators at randomisation: patients might not be as treatment resistant as in other trials
	Allocation concealment		X		
	Blinding (participants and personnel)		X		
	Blinding of outcomes assessment		X		
	Incomplete outcome data		X		
	Selective reporting		X		
	'Other bias'		X		
Mean change in MADRS	Random sequence generation		X		Limited details reported on trial methodology. Potential issue of non-blinding of investigators at randomisation: patients might not be as treatment resistant as in other trials
	Allocation concealment		X		
	Blinding (participants and personnel)		X		
	Blinding of outcomes assessment		X		
	Incomplete outcome data		X		
	Selective reporting		X		
	'Other bias'		X		

Outcome	Risk of bias	Low risk	Unclear risk	High risk	Comments
Remission	Random sequence generation		X		Limited details reported on trial methodology. Potential issue of non-blinding of investigators at randomisation: patients might not be as treatment resistant as in other trials
	Allocation concealment		X		
	Blinding (participants and personnel)		X		
	Blinding of outcomes assessment		X		
	Incomplete outcome data		X		
	Selective reporting		X		
	'Other bias'		X		
Withdrawals (all cause)	Random sequence generation		X		Limited details reported on trial methodology. Potential issue of non-blinding of investigators at randomisation: patients might not be as treatment resistant as in other trials
	Allocation concealment		X		
	Blinding (participants and personnel)		X		
	Blinding of outcomes assessment		X		
	Incomplete outcome data		X		
	Selective reporting		X		
	'Other bias'		X		
Adverse events	Random sequence generation		X		Limited details reported on trial methodology. Potential issue of non-blinding of investigators at randomisation: patients might not be as treatment resistant as in other trials
	Allocation concealment		X		
	Blinding (participants and personnel)		X		
	Blinding of outcomes assessment		X		
	Incomplete outcome data		X		
	Selective reporting		X		
	'Other bias'		X		
Overall trial risk-of-bias rating:					
Low risk	Unclear			High risk	

Thase *et al.* 2007⁵³ (studies 'a' and 'b')

Outcome	Risk of bias	Low risk	Unclear risk	High risk	Comments
Response	Random sequence generation		X		Limited details reported on trial methodology
	Allocation concealment		X		
	Blinding (participants and personnel)		X		
	Blinding of outcomes assessment		X		
	Incomplete outcome data		X		
	Selective reporting		X		
	'Other bias'		X		

Outcome	Risk of bias	Low risk	Unclear risk	High risk	Comments
Mean change in MADRS	Random sequence generation		X		Limited details reported on trial methodology
	Allocation concealment		X		
	Blinding (participants and personnel)		X		
	Blinding of outcomes assessment		X		
	Incomplete outcome data		X		
	Selective reporting		X		
	'Other bias'		X		
Remission	Random sequence generation		X		Limited details reported on trial methodology
	Allocation concealment		X		
	Blinding (participants and personnel)		X		
	Blinding of outcomes assessment		X		
	Incomplete outcome data		X		
	Selective reporting		X		
	'Other bias'		X		
Withdrawals (all cause)	Random sequence generation		X		Limited details reported on trial methodology
	Allocation concealment		X		
	Blinding (participants and personnel)		X		
	Blinding of outcomes assessment		X		
	Incomplete outcome data	X			
	Selective reporting		X		
	'Other bias'		X		
Adverse events	Random sequence generation		X		Limited details reported on trial methodology
	Allocation concealment		X		
	Blinding (participants and personnel)		X		
	Blinding of outcomes assessment		X		
	Incomplete outcome data		X		
	Selective reporting		X		
	'Other bias'		X		
QoL	Random sequence generation		X		Limited details reported on trial methodology
	Allocation concealment		X		
	Blinding (participants and personnel)		X		
	Blinding of outcomes assessment		X		
	Incomplete outcome data		X		
	Selective reporting		X		
	'Other bias'		X		
Overall trial risk-of-bias rating:					
Low risk	Unclear		High risk		

Appendix 6 Summary of trials included for each outcome in the network meta-analysis

Study	Outcome												
	Response (using lithium a priori data) ^a	Response (using lithium post hoc data) ^{a,b}	Mean change in MADRS score from baseline ^{a,c}	Withdrawals (all cause) ^d	Total adverse events ^d	Somnolence ^a	Dry mouth ^a	Headache ^a	SA 1 ^a	SA 2 ^a	SA 3 ^a	SA 4 ^a	SA 5 ^d
Berman <i>et al.</i> (2007) ⁴⁷	X	X	X	X	X	X	X	X	✓	X	X	✓	✓
Berman <i>et al.</i> (2009) ⁴⁸	X	X	X	X	X	X	X	X	✓	X	X	✓	✓
Corya <i>et al.</i> (2006) ⁴⁹	✓	✓	✓	✓	X	✓	✓	✓	✓	X	✓	✓	✓
Feng <i>et al.</i> (2008) ⁴³	✓	✓	X	X	✓	X	X	X	✓	X	X	X	X
Franco <i>et al.</i> (2010) ³⁰	X	X	X	X	X	X	X	X	✓	X	X	✓	X
Katona <i>et al.</i> (1995) ⁵⁹	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Marcus <i>et al.</i> (2008) ⁵⁰	X	X	X	X	X	X	X	X	✓	X	X	✓	✓
Mattingly <i>et al.</i> (2006) ⁴⁶	X	X	X	X	X	X	X	X	✓	X	X	X	✓
Shelton <i>et al.</i> (2001) ⁵¹	✓	✓	X	✓	X	X	X	X	✓	✓	✓	✓	✓
Shelton <i>et al.</i> (2005) ⁵²	✓	✓	✓	✓	✓	X	X	X	✓	X	✓	✓	✓
Thase <i>et al.</i> (2007a) ⁵³	✓	✓	✓	✓ ^e	X	✓ ^e	✓ ^e	✓ ^e	✓	✓	✓	✓	✓
Thase <i>et al.</i> (2007b) ⁵³	✓	✓	✓	✓ ^e	X	✓ ^e	✓ ^e	✓ ^e	✓	✓	✓	✓	✓

a Random-effects model.
 b Primary analysis for this review.
 c (Weighted) MD.
 d Fixed-effects model.
 e Thase *et al.*⁵³ (a+b) pooled study data used for this outcome.
 Notes
 SA 1 class-based sensitivity analysis for response.
 SA 2 sensitivity analysis for response limiting to trials reporting failure to two or more antidepressants in the current episode of depression.
 SA 3 sensitivity analysis for response limiting to trials reporting response based on the MADRS scale.
 SA 4 class-based sensitivity analysis for mean change in MADRS.
 SA 5 class-based sensitivity analysis for discontinuations.

Appendix 7 Health economics literature search strategies

Economics search strategy

Database: EMBASE

Search strategy

1. *depression/ or *major depression/
2. depress\$.tw.
3. or/1-2
4. (refract\$ or resistan\$ or nonrespon\$ or unrespon\$ or fail\$ or (incomplet\$ adj respon\$) or (no\$ adj2 respon\$)).tw.
5. drug treatment failure/
6. (inadequat\$ respon\$ or (sub\$ adj2 respon\$) or (poor\$ adj respon\$)).tw.
7. or/4-6
8. (anti-depress\$ or antidepress\$).tw.
9. antidepressant agent/
10. 8 or 9
11. (augment\$ or adjunct\$ or combin\$ or add\$ or potentiation).tw.
12. drug combination/
13. 11 or 12
14. exp lithium/ or exp lithium derivative/ or lithium carbonate/ or lithium chloride/ or lithium citrate/
15. atypical antipsychotic agent/ or ziprasidone/ or risperidone/ or quetiapine/ or paliperidone/ or olanzapine/ or clozapine/ or clozapine derivative/ or clozapine n oxide/ or aripiprazole/ or amisulpride/
16. (Lithium or Camcolit or Liskonum or Priadel or Li-Liquid or Litarex or Lithonate or Phasal or Lithny or Licio or Lithii or Litio or Litium or Lityum or dilithium or litu or Cibalith-S or Eskalith or Lithane or Lithobid or Lithonate or Lithotabs).tw.
17. ((atypical adj antipsychotic?) or (atypical adj anti-psychotic?) or Amisulprid\$ or Solian or Aripiprazol\$ or Abilify or Clozapin\$ or Clozaril or Klotsapiini or Klozapin\$ or FazaClo or Fazalco or Denzapine or Zaponex or Olanzapin\$ or Olantsapiini or Zyprexa or Zypadhera or Paliperidon\$ or Invega or Xeplion or Quetiapin\$ or Seroquel or Risperidon\$ or Riszperidon or Rysperydin or Risperdal or Ziprasidon\$ or Geodon).tw.
18. or/14-17
19. serotonin antagonist/ or serotonin uptake inhibitor/ or fluoxetine plus olanzapine/ or fluvoxamine maleate/ or ssri\$.tw. or (serotonin adj2 inhibitor\$).tw.
20. (citalopram or citalopramum or celexa or cipramil).tw. or citalopram/
21. (escitalopram or S-citalopram or lexapro or cipralox).tw. or escitalopram/
22. (fluoxetin\$ or fluokset\$ or feliciam or prozac or prozep or prozit or sarafem or selfemra or symbyax).tw. or fluoxetine/ or fluoxetine plus olanzapine/
23. (fluvoxamin\$ or fluvoksami\$ or faverin or luvox).tw. or fluvoxamine/ or fluvoxamine maleate/
24. (paroxetin\$ or parokset\$ or seroxat or paxil or pexeva).tw. or paroxetine/
25. (sertralin\$ or lustral or zolof).tw. or sertraline/
26. or/19-25
27. 3 and 7
28. 10 and 13 and 18
29. 26 or 28
30. 27 and 29
31. health economics/

32. exp economic evaluation/
33. exp pharmacoeconomics/
34. exp health care cost/
35. or/31-34
36. (price or prices or pricing).tw.
37. value for money.tw.
38. (economic\$ or pharmaeconomic\$ or pharmacoeconomic\$ or pharmaco-economic\$).tw.
39. (cost or costs or costly or costing or costed).tw.
40. or/36-39
41. 35 or 40
42. letter.pt.
43. editorial.pt.
44. note.pt.
45. or/42-44
46. 41 not 45
47. 30 and 46

Database: MEDLINE

Search strategy

1. Depression/ or Depressive Disorder/ or Depressive Disorder, Major/
2. depress\$.tw.
3. or/1-2
4. treatment failure/
5. (inadequat\$ respon\$ or (sub\$ adj2 respon\$) or (poor\$ adj respon\$)).tw.
6. (refract\$ or resistanc\$ or nonrespon\$ or unrespon\$ or fail\$ or (incomplet\$ adj respon\$) or (no\$ adj2 respon\$)).tw.
7. or/4-6
8. 3 and 7
9. (anti-depress\$ or antidepress\$).tw.
10. Antidepressive Agents/
11. 9 or 10
12. (augment\$ or adjunct\$ or combin\$ or add\$ or potentiation).tw.
13. Drug Therapy, Combination/
14. 12 or 13
15. Serotonin Uptake Inhibitors/ or Antidepressive Agents, Second-Generation/ or ssri\$.tw. or (serotonin adj2 inhibitor\$).tw. or Serotonin Antagonists/
16. (citalopram or citalopramum or celexa or cipramil).tw. or Citalopram/
17. (escitalopram or S-citalopram or lexapro or cipralext).tw.
18. (fluoxetin\$ or fluokset\$ or feliciu\$ or prozac or prozep or prozite\$ or sarafem or selfemra or symbyax).tw. or Fluoxetine/
19. (fluvoxamin\$ or fluvoksami\$ or faverin or luvox).tw. or Fluvoxamine/
20. (paroxetin\$ or parokset\$ or seroxat or paxil or pexeva).tw. or Paroxetine/
21. (sertralin\$ or lustral or zoloft).tw. or Sertraline/
22. or/15-21
23. Lithium/ or exp Lithium Compounds/
24. antipsychotic agents/ or clozapine/ or risperidone/
25. (Lithium or Camcolit or Liskonum or Priadel or Li-Liquid or Litarex or Lithonate or Phasal or Lithny or Licio or Lithii or Litio or Litium or Lityum or dilithium or litu or Cibalith-S or Eskalith or Lithane or Lithobid or Lithonate or Lithotabs).tw.

26. ((atypical adj antipsychotic?) or (atypical adj anti-psychotic?) or Amisulprid\$ or Solian or Aripiprazol\$ or Abilify or Clozapin\$ or Clozaril or Klotsapiini or Klozapin\$ or FazaClo or Fazalco or Denzapine or Zaponex or Olanzapin\$ or Olantsapiini or Zyprexa or Zypadhera or Paliperidon\$ or Invega or Xeplion or Quetiapin\$ or Seroquel or Risperidon\$ or Riszperidon or Rysperydol or Risperdal or Ziprasidon\$ or Geodon).tw.
27. or/23-26
28. 11 and 14 and 27
29. 22 or 28
30. 8 and 29
31. economics/
32. exp costs/ and cost analysis/
33. exp economics, hospital/
34. economics, medical/
35. economics, pharmaceutical/
36. (economic\$ or pharmaeconomic\$ or pharmacoeconomic\$ or pharmaco-economic\$).tw.
37. (cost or costs or costly or costing or costed).tw.
38. (price or prices or pricing).tw.
39. value for money.tw.
40. or/31-39
41. letter.pt.
42. editorial.pt.
43. comment.pt.
44. or/41-43
45. 40 not 44
46. 30 and 45

Database: PsycINFO

Search strategy

1. costs.mp. and cost analysis/ [mp=title, abstract, heading word, table of contents, key concepts, original title, tests & measures]
2. cost containment/
3. pharmacoeconomics/
4. health care economics/
5. or/1-4
6. (economic adj2 evaluation\$).tw.
7. (economic adj2 analy\$).tw.
8. (economic adj2 (study or studies)).tw.
9. (cost adj2 evaluation\$).tw.
10. (cost adj2 analy\$).tw.
11. (cost adj2 (study or studies)).tw.
12. (cost adj2 effective\$).tw.
13. (cost adj2 benefit\$).tw.
14. (cost adj2 utili\$).tw.
15. (cost adj2 minimi\$).tw.
16. (cost adj2 consequence\$).tw.
17. (cost adj2 comparison\$).tw.
18. (cost adj2 identificat\$).tw.
19. (pharmaeconomic\$ or pharmacoeconomic\$ or pharmaco-economic\$).tw.
20. or/6-19
21. 5 or 20
22. editorial.dt.

23. letter.dt.
24. or/22-23
25. 21 not 24
26. exp Major Depression/
27. depress\$.tw.
28. or/26-27
29. (refract\$ or resistan\$ or nonrespon\$ or unrespon\$ or fail\$ or (incomplet\$ adj respon\$) or (no\$ adj2 respon\$)).tw.
30. (inadequat\$ respon\$ or (sub\$ adj2 respon\$) or (poor\$ adj respon\$)).tw.
31. or/29-30
32. antidepressant drugs/
33. (antidepress\$ or anti-depress\$).tw.
34. or/32-33
35. (augment\$ or adjunct\$ or combin\$ or add\$ or potentiation).tw.
36. polypharmacy/
37. Drug Augmentation/
38. or/35-37
39. Serotonin Antagonists/ or serotonin reuptake inhibitors/ or (ssri\$ or (serotonin adj2 inhibitor\$)).tw.
40. (citalopram or citalopramum or celexa or cipramil).tw. or Citalopram/
41. (escitalopram or S-citalopram or lexapro or cipralext).tw.
42. (fluoxetin\$ or fluokset\$ or feliciam or prozac or prozep or prozit or sarafem or selfemra or symbyax).tw.
43. (fluvoxamin\$ or fluvoksami\$ or faverin or luvox).tw. or Fluvoxamine/
44. (paroxetin\$ or parokset\$ or seroxat or paxil or pexeva).tw. or Paroxetine/
45. or/39-44
46. (Lithium or Camcolit or Liskonum or Priadel or Li-Liquid or Litarex or Lithonate or Phasal or Lithny or Licio or Lithii or Litio or Litium or Lityum or dilithium or litu or Cibalith-S or Eskalith or Lithane or Lithobid or Lithonate or Lithotabs).tw.
47. exp Lithium Carbonate/ or Lithium/
48. neuroleptic drugs/ or aripiprazole/ or clozapine/ or olanzapine/ or quetiapine/ or risperidone/
49. ((atypical adj antipsychotic?) or (atypical adj anti-psychotic?) or Amisulprid\$ or Solian or Aripiprazol\$ or Abilify or Clozapin\$ or Clozaril or Klotsapiini or Klozapin\$ or FazaClo or Fazalco or Denzapine or Zaponex or Olanzapin\$ or Olantsapiini or Zyprexa or Zypadhera or Paliperidon\$ or Invega or Xeplion or Quetiapin\$ or Seroquel or Risperidon\$ or Risperidon or Risperidon or Risperidon or Risperidon or Risperidon\$ or Ziprasidon\$ or Geodon).tw.
50. or/46-49
51. 28 and 31
52. 34 and 38 and 50
53. 45 or 52
54. 51 and 53
55. 25 and 54

Quality-of-life search strategies

Database: EMBASE

Quality-of-life search strategy

1. exp *quality of life/
2. quality of life.tw.
3. (quality adjusted life year\$ or quality-adjusted life year\$).tw.
4. qaly\$.tw.
5. qol.tw.

6. hrqol.tw.
7. euroqol.tw.
8. short form 36.tw.
9. (disability adjusted life year\$ or disability-adjusted life year\$).tw.
10. ((value or valuation or quality) adj2 (life or lives or survival)).tw.
11. or/1-10
12. letter.pt.
13. editorial.pt.
14. note.pt.
15. or/12-14
16. 11 not 15
17. *depression/ or *major depression/
18. depress\$.tw.
19. or/17-18
20. (refract\$ or resistan\$ or nonrespon\$ or unrespon\$ or fail\$ or (incomplet\$ adj respon\$) or (no\$ adj2 respon\$)).tw.
21. drug treatment failure/
22. (inadequat\$ respon\$ or (sub\$ adj2 respon\$) or (poor\$ adj respon\$)).tw.
23. or/20-22
24. (anti-depress\$ or antidepress\$).tw.
25. antidepressant agent/
26. 24 or 25
27. (augment\$ or adjunct\$ or combin\$ or add\$ or potentiation).tw.
28. drug combination/
29. 27 or 28
30. exp lithium/ or exp lithium derivative/ or lithium carbonate/ or lithium chloride/ or lithium citrate/
31. atypical antipsychotic agent/ or ziprasidone/ or risperidone/ or quetiapine/ or paliperidone/ or olanzapine/ or clozapine/ or clozapine derivative/ or clozapine n oxide/ or aripiprazole/ or amisulpride/
32. (Lithium or Camcolit or Liskonum or Priadel or Li-Liquid or Litarex or Lithonate or Phasal or Lithny or Licio or Lithii or Litio or Litium or Lityum or dilithium or litu or Cibalith-S or Eskalith or Lithane or Lithobid or Lithonate or Lithotabs).tw.
33. ((atypical adj antipsychotic?) or (atypical adj anti-psychotic?) or Amisulprid\$ or Solian or Aripiprazol\$ or Abilify or Clozapin\$ or Clozaril or Klotsapiini or Klozapin\$ or FazaClo or Fazalco or Denzapine or Zaponex or Olanzapin\$ or Olantsapiini or Zyprexa or Zypadhera or Paliperidon\$ or Invega or Xeplion or Quetiapin\$ or Seroquel or Risperidon\$ or Riszperidon or Rysperidon or Risperdal or Ziprasidon\$ or Geodon).tw.
34. or/30-33
35. serotonin antagonist/ or serotonin uptake inhibitor/ or fluoxetine plus olanzapine/ or fluvoxamine maleate/ or ssri\$.tw. or (serotonin adj2 inhibitor\$).tw.
36. (citalopram or citalopramum or celexa or cipramil).tw. or citalopram/
37. (escitalopram or S-citalopram or lexapro or cipralex).tw. or escitalopram/
38. (fluoxetin\$ or fluokset\$ or felicium or prozac or prozep or prozit or sarafem or selfemra or symbyax).tw. or fluoxetine/ or fluoxetine plus olanzapine/
39. (fluoxamin\$ or fluvoksami\$ or faverin or luvox).tw. or fluvoxamine/ or fluvoxamine maleate/
40. (paroxetin\$ or parokset\$ or seroxat or paxil or pexeva).tw. or paroxetine/
41. (sertralin\$ or lustral or zoloft).tw. or sertraline/
42. or/35-41
43. 19 and 23
44. 26 and 29 and 34
45. 42 or 44
46. 43 and 45
47. 16 and 46

Database: MEDLINE**Quality-of-life search strategy**

1. quality of life/
2. value of life/
3. quality-adjusted life years/
4. cost of illness/
5. quality of life.tw.
6. (quality adjusted life year\$ or quality-adjusted life year\$).tw.
7. qaly\$.tw.
8. qol.tw.
9. hrqol.tw.
10. euroqol.tw.
11. short form 36.tw.
12. (disability adjusted life year\$ or disability-adjusted life year\$).tw.
13. ((value or valuation or quality) adj2 (life or lives or survival)).tw.
14. or/1-4
15. or/5-13
16. 14 or 15
17. letter.pt.
18. editorial.pt.
19. comment.pt.
20. or/17-19
21. 16 not 20
22. Depression/ or Depressive Disorder/ or Depressive Disorder, Major/
23. depress\$.tw.
24. or/22-23
25. treatment failure/
26. (inadequat\$ respon\$ or (sub\$ adj2 respon\$) or (poor\$ adj respon\$)).tw.
27. (refract\$ or resistan\$ or nonrespon\$ or unrespon\$ or fail\$ or (incomplet\$ adj respon\$) or (no\$ adj2 respon\$)).tw.
28. or/25-27
29. 24 and 28
30. (anti-depress\$ or antidepress\$).tw.
31. Antidepressive Agents/
32. 30 or 31
33. (augment\$ or adjunct\$ or combin\$ or add\$ or potentiation).tw.
34. Drug Therapy, Combination/
35. 33 or 34
36. Serotonin Uptake Inhibitors/ or Antidepressive Agents, Second-Generation/ or ssri\$.tw. or (serotonin adj2 inhibitor\$).tw. or Serotonin Antagonists/
37. (citalopram or citalopramum or celexa or cipramil).tw. or Citalopram/
38. (escitalopram or S-citalopram or lexapro or cipralax).tw.
39. (fluoxetin\$ or fluokset\$ or feliciem or prozac or prozep or prozit or sarafem or selfemra or symbyax).tw. or Fluoxetine/
40. (fluvoxamin\$ or fluvoksami\$ or faverin or luvox).tw. or Fluvoxamine/
41. (paroxetin\$ or parokset\$ or seroxat or paxil or pexeva).tw. or Paroxetine/
42. (sertralin\$ or lustral or zoloft).tw. or Sertraline/
43. or/36-42
44. Lithium/ or exp Lithium Compounds/
45. antipsychotic agents/ or clozapine/ or risperidone/

46. (Lithium or Camcolit or Liskonum or Priadel or Li-Liquid or Litarex or Lithonate or Phasal or Lithny or Licio or Lithii or Litio or Litium or Lityum or dilithium or litu or Cibalith-S or Eskalith or Lithane or Lithobid or Lithonate or Lithotabs).tw.
47. ((atypical adj antipsychotic?) or (atypical adj anti-psychotic?) or Amisulprid\$ or Solian or Aripiprazol\$ or Abilify or Clozapin\$ or Clozaril or Klotsapiini or Klozapin\$ or FazaClo or Fazalco or Denzapine or Zaponex or Olanzapin\$ or Olantsapiini or Zyprexa or Zypadhera or Paliperidon\$ or Invega or Xeplion or Quetiapin\$ or Seroquel or Risperidon\$ or Riszperidon or Rysperydol or Risperdal or Ziprasidon\$ or Geodon).tw.
48. or/44-47
49. 32 and 35 and 48
50. 43 or 49
51. 29 and 50
52. 21 and 51

Database: PsycINFO

Quality-of-life search strategy

1. quality of life/
2. quality of life.tw.
3. (quality adjusted life year\$ or quality-adjusted life year\$).tw.
4. qaly\$.tw.
5. qol.tw.
6. hrqol.tw.
7. euroqol.tw.
8. short form 36.tw.
9. (disability adjusted life year\$ or disability-adjusted life year\$).tw.
10. ((value or valuation or quality) adj2 (life or lives or survival)).tw.
11. or/1-10
12. editorial.dt.
13. letter.dt.
14. or/12-13
15. 11 not 14
16. exp Major Depression/
17. depress\$.tw.
18. or/16-17
19. (refract\$ or resistan\$ or nonrespon\$ or unrespon\$ or fail\$ or (incomplet\$ adj respon\$) or (no\$ adj2 respon\$)).tw.
20. (inadequat\$ respon\$ or (sub\$ adj2 respon\$) or (poor\$ adj respon\$)).tw.
21. or/19-20
22. antidepressant drugs/
23. (antidepress\$ or anti-depress\$).tw.
24. or/22-23
25. (augment\$ or adjunct\$ or combin\$ or add\$ or potentiation).tw.
26. polypharmacy/
27. Drug Augmentation/
28. or/25-27
29. Serotonin Antagonists/ or serotonin reuptake inhibitors/ or (ssri\$ or (serotonin adj2 inhibitor\$)).tw.
30. (citalopram or citalopramum or celexa or cipramil).tw. or Citalopram/
31. (escitalopram or S-citalopram or lexapro or cipralox).tw.
32. (fluoxetin\$ or fluokset\$ or feliciam or prozac or prozep or prozit or sarafem or selfemra or symbyax).tw.
33. (fluvoxamin\$ or fluvoxami\$ or faverin or luvox).tw. or Fluvoxamine/
34. (paroxetin\$ or parokset\$ or seroxat or paxil or pexeva).tw. or Paroxetine/

35. or/29-34
36. (Lithium or Camcolit or Liskonum or Priadel or Li-Liquid or Litarex or Lithonate or Phasal or Lithny or Licio or Lithii or Lio or Litium or Lityum or dilithium or litu or Cibalith-S or Eskalith or Lithane or Lithobid or Lithonate or Lithotabs).tw.
37. exp Lithium Carbonate/ or Lithium/
38. neuroleptic drugs/ or aripiprazole/ or clozapine/ or olanzapine/ or quetiapine/ or risperidone/
39. ((atypical adj antipsychotic?) or (atypical adj anti-psychotic?) or Amisulprid\$ or Solian or Aripiprazol\$ or Abilify or Clozapin\$ or Clozaril or Klotsapiini or Klozapin\$ or FazaClo or Fazalco or Denzapine or Zaponex or Olanzapin\$ or Olantsapiini or Zyprexa or Zypadhera or Paliperidon\$ or Invega or Xeplion or Quetiapin\$ or Seroquel or Risperidon\$ or Riszperidon or Rysperydol or Risperdal or Ziprasidon\$ or Geodon).tw.
40. or/36-39
41. 18 and 21
42. 24 and 28 and 40
43. 35 or 42
44. 41 and 43
45. 15 and 44

Database: NHS Economic Evaluation Database, Health Technology Assessment Database

Search strategy

- #1 Depression
- #2 Depressive Disorder
- #3 Mood Disorders
- #4 (depress* or dysphori* or dysthym* or seasonal affective disorder* or melanchol*):
ti or (depress* or dysphori* or dysthym* or seasonal affective disorder*
or melanchol*):ab
- #5 (#1 OR #2 OR #3 OR #4)
- #6 lithium
- #7 atypical antipsychotics
- #8 serotonin reuptake inhibitors OR SSRI
- #9 (#6 OR #7 OR #8)
- #10 (#5 AND #9)

Database: The Cochrane Library

Search strategy

- #1 Depressive Disorder in Economic Evaluations
- #2 Depression in Economic Evaluations
- #3 depress*:ti,ab in Economic Evaluations

- #4 Major Depressive Disorder in Economic Evaluations
- #5 (#1 OR #2 OR #3 OR #4)
- #6 Antidepressive Agents OR Second-Generation in Economic Evaluations
- #7 Serotonin Antagonists in Economic Evaluations
- #8 Serotonin Uptake Inhibitors in Economic Evaluations
- #9 Citalopram in Economic Evaluations
- #10 Fluoxetine in Economic Evaluations
- #11 Fluvoxamine in Economic Evaluations
- #12 Paroxetine in Economic Evaluations
- #13 Sertraline in Economic Evaluations
- #14 ssri* or (serotonin near/2 inhibitor*) or citalopram or citalopramum or celexa or cipramil or escitalopram or S-citalopram or lexapro or cipralex or fluoxetin* or fluokset* or felicium or prozac or prozep or prozit or sarafem or selfemra or symbyax or fluvoxamin* or fluvoksami* or faverin or luvox or paroxetin* or parokset* or seroxat or paxil or pexeva or sertralin* or lustral or zoloft in Economic Evaluations
- #15 (#6 OR #7 OR #8 OR #9 OR #10 OR #11 OR #12 OR #13 OR #14)
- #16 refract* or resistan* or nonrespon* or (no* near/2 respon*) or unrespon* or fail* or (incomplet* near/1 respon*) or (inadequat* near/1 respons*) or (sub* near/2 respon*) or (poor* near/1 respon*): ti,ab in Economic Evaluations
- #17 Treatment Failure in Economic Evaluations
- #18 (#16 OR #17)
- #19 antidepress* or anti-depress*:ti,ab in Economic Evaluations
- #20 Antidepressive Agents in Economic Evaluations
- #21 (#19 OR #20)
- #22 augment* or adjunct* or combin* or add* or potentiation:ti,ab in Economic Evaluations
- #23 Drug Therapy in Economic Evaluations
- #24 (#22 OR #23)
- #25 Lithium or Camcolit or Liskonum or Priadel or Li-Liquid or Litarex or Lithonate or Phasal or Lithny or Licio or Lithii or Litio or Litium or Lityum or dilithium or litu or Cibalith-S or Eskalith or Lithane or Lithobid or Lithonate or Lithotabs in Economic Evaluations
- #26 (atypical near/1 antipsychotic*) or (atypical near/1 anti-psychotic*) or Amisulprid* or Solian or Aripiprazol* or Abilify or Clozapin* or Clozaril or Klotsapiini or Klozapin* or FazaClo or Fazalco or Denzapine or Zaponex or Olanzapin* or Olantsapiini or Zyprexa or Zypadhera or Paliperidon* or Invega or Xeplion or Quetiapin* or Seroquel or Risperidon* or Riszperidon or Rysperydol or Risperdal or Ziprasidon* or Geodon in Economic Evaluations
- #27 Lithium in Economic Evaluations
- #28 Lithium Compounds in Economic Evaluations
- #29 Antipsychotic Agents in Economic Evaluations

- #30 Clozapine in Economic Evaluations
- #31 Risperidone in Economic Evaluations
- #32 (#25 OR #26 OR #27 OR #28 OR #29 OR #30 OR #31)
- #33 (#5 AND 15 AND #18 AND #21 AND #24 AND #32)

Appendix 8 Quality assessment of included studies (cost-effectiveness studies)

Leelahanj 2010⁸⁸

Dimension of quality	Comments
Structure	
S1: Statement of decision problem/objective	Clearly stated
S2: Statement of scope/perspective	The model scope and perspective (Thai) are clearly stated and the model outcomes are consistent with the scope and overall objective of the model
S3: Rationale for structure	Although no justification was given for adopting a decision tree, the author's approach is considered appropriate given that the model focused on acute treatment of MDD. The model was well constructed and data sources were well described
S4: Structural assumptions	The model is considered to be well constructed and assumptions clearly stated. However, it is important to note that authors made some very optimistic assumptions that may not be credible. For instance adverse events were not modelled and there was no justification as to why they did not include them. Also the assumption that all patients who are given ECT will enter remission may not necessarily be true
S5: Strategies/comparators	The model compared adjunctive therapy with aripiprazole or placebo
S6: Model type	Correct, cost–utility analysis
S7: Time horizon	Six weeks matching the duration of trial data
S8: Disease states/pathways	The pathways/health states modelled are relevant. However, authors could have also included response rates as a separate health state as this is one of the most common reported outcomes in MDD
S9: Cycle length	This was a decision tree that captured acute treatment outcomes
Data	
D1: Data identification	Data were systematically sourced, clearly described and justified by the authors
D2: Pre-model data analysis	The authors did not report how they synthesised the trial data they used in the model. Data derived from the trials were directly implemented in the model
D2a: Baseline data	Events reported in either placebo or aripiprazole were used to populate the decision tree
D2b: Treatment effects	Events reported in either placebo or aripiprazole were used to populate the decision tree. RRs and 95% CIs were calculated for aripiprazole
D2c: Costs	Appropriate costs were included and source of cost data clearly described
D2d: Quality-of-life weights (utilities)	Derived from literature and clearly referenced
D3: Data incorporation	The authors clearly described how data were used in the model
D4: Assessment of uncertainty	The assessment of sensitivity was thorough and robust
D4a: Methodological	The authors used appropriate analytical methods for the decision problem
D4b: Structural	The authors described deterministic sensitivity analysis and scenario analysis undertaken for the model
D4c: Heterogeneity	Heterogeneity was not addressed, no subgroup analysis was undertaken
D4d: Parameter	OWSA and scenario analysis were undertaken. PSA was not carried out

Dimension of quality	Comments
Consistency	
C1: Internal consistency	The model seems to be mathematically sound with no obvious inconsistencies
C2: External consistency	The results of the model are applicable to patients with MDD. Efficacy data were taken from two US studies that may not be directly applicable to the UK population. Costs data were taken from local Thai hospital, which has a different health-care system to the UK. Overall the study findings and conclusions cannot be generalised to the UK setting

Xie 2009⁸⁷

Dimension of quality	Comments
Structure	
S1: Statement of decision problem/objective	Clearly stated
S2: Statement of scope/perspective	The model scope and perspective are clearly stated and the model outcomes are consistent with the scope and overall objective of the model. The analysis was conducted from a Singaporean societal perspective
S3: Rationale for structure	Although no justification was given for adopting a decision tree, the authors' approach is considered appropriate and is in line with other published economic models in the area of MDD. The model was well constructed and data sources were well described
S4: Structural assumptions	The model is considered to be well constructed and assumptions clearly stated. However, it is important to note that authors made some very optimistic assumptions that may not be credible. Owing to lack of head-to-head trial of escitalopram vs. fluvoxamine, authors used efficacy data for citalopram as a proxy for the efficacy of fluvoxamine. Also the assumption that relapse rates following augmentation or combination therapy are the same following first-line treatment contradicts external evidence
S5: Strategies/comparators	The model compared escitalopram vs. venlafaxine and vs. fluvoxamine
S6: Model type	Cost-effectiveness analysis
S7: Time horizon	Six months following the initiation of therapy, in accordance with the Singaporean depression clinical guideline
S8: Disease states/pathways	The pathways/health states modelled are relevant. However, authors could have also included response rates as a separate health state as this is one of the most common reported outcomes in MDD
S9: Cycle length	This was a decision tree which captured 6 months' treatment outcomes
Data	
D1: Data identification	Data were systematically sourced, clearly described and justified by the authors
D2: Pre-model data analysis	The authors did not report any pre-model data analysis
D2a: Baseline data	Events reported in the included studies were used in the model. Remission rates were taken from a referenced meta-analysis and as stated earlier data for citalopram were used as a proxy for the efficacy of fluvoxamine
D2b: Treatment effects	Events reported in either escitalopram vs. venlafaxine or escitalopram vs. fluvoxamine and 95% CrIs were reported
D2c: Costs	Costs appropriate to the perspective of the analysis were used, sources were clearly stated
D2d: Quality-of-life weights (utilities)	Not reported, this was not a cost-utility analysis
D3: Data incorporation	The authors clearly described how data were used in the model
D4: Assessment of uncertainty	The assessment of sensitivity was thorough and robust as both deterministic and PSA were reported

Dimension of quality	Comments
D4a: Methodological	The authors used appropriate analytical methods for the decision problem; however, they could have undertaken a cost–utility study
D4b: Structural	The authors described deterministic sensitivity analysis and PSA was undertaken for the model
D4c: Heterogeneity	Heterogeneity was partially addressed as the model considered primary-care patients and secondary care in separate analysis. However, there was no further subgroup analysis reported
D4d: Parameter	Both deterministic and PSA were reported
Consistency	
C1: Internal consistency	The model seems to be mathematically sound with no obvious inconsistencies
C2: External consistency	The results of the model are applicable to patients with MDD. Efficacy data were taken from studies conducted in the US, which may not be directly applicable to the UK population. Costs data were estimated from a survey of local practitioners in Singapore, which has a different health-care system to the UK. Overall, the study findings and conclusions cannot be generalised to the UK setting

Benedict 2010⁸⁶

Dimension of quality	Comments
Structure	
S1: Statement of decision problem/objective	Clearly stated
S2: Statement of scope/perspective	The model scope and perspective are clearly stated and the model outcomes are consistent with the scope and overall objective of the model. The model was conducted from a Scottish NHS perspective
S3: Rationale for structure	Although no justification was given for adopting a Markov, the authors' approach is considered appropriate given the cyclical nature of the disease
S4: Structural assumptions	The model is considered to be well constructed and assumptions clearly stated
S5: Strategies/comparators	The model compared duloxetine vs. venlafaxine ER or mirtazapine
S6: Model type	Correct, cost–utility analysis
S7: Time horizon	Forty-eight-week time horizon chosen to capture treatment duration as recommended by NICE [6 months following remission (NICE 2007)] and relapses within 1 year
S8: Disease states/pathways	The pathways/health states modelled are relevant
S9: Cycle length	Eight-week cycle length was chosen, which is the typical treatment duration of most trials in MDD

Dimension of quality	Comments
Data	
D1: Data identification	Data were systematically sourced, clearly described and justified by the authors
D2: Pre-model data analysis	The authors did not report any pre-model analysis. Data derived from different meta-analysis for different parameters was directly implemented in the model
D2a: Baseline data	Events reported in different meta-analysis studies were used to populate the model
D2b: Treatment effects	Treatment effect reported in different meta-analysis studies were used to populate the model
D2c: Costs	Appropriate costs included in base case and sensitivity analysis, sources are transparent
D2d: Quality-of-life weights (utilities)	Derived from literature and clearly referenced
D3: Data incorporation	The authors clearly described how data were used in the model
D4: Assessment of uncertainty	The assessment of sensitivity was thorough and robust as both deterministic and PSA were undertaken
D4a: Methodological	The authors used appropriate analytical methods for the decision problem
D4b: Structural	The authors described deterministic sensitivity analysis and PSA was also undertaken
D4c: Heterogeneity	Heterogeneity was partially assessed by reporting results separately for primary and secondary-care patients; however, no further subgroup analysis was undertaken
D4d: Parameter	OWSA and PSA were undertaken
Consistency	
C1: Internal consistency	The model seems to be mathematically sound with no obvious inconsistencies
C2: External consistency	The results of the model are applicable to patients with MDD. Resource-use data were provided by a panel of practising practitioners in Scotland and reflected the Scottish practice. Overall, the study findings and conclusions cannot be directly generalised to the UK setting but are nonetheless informative

Simpson 2009⁸⁹

Dimension of quality	Comments
Structure	
S1: Statement of decision problem/objective	Clearly stated
S2: Statement of scope/perspective	The model scope and perspective are clearly stated and the model outcomes are consistent with the scope and overall objective of the model. The model was conducted from a US payer and societal perspective
S3: Rationale for structure	No justification was given for adopting a hybrid model structure, i.e. a combination of decision tree for the acute phase of the model and then a Markov for the maintenance phase of the model. However, this approach is considered appropriate
S4: Structural assumptions	The model is considered to be well constructed and assumptions clearly stated
S5: Strategies/comparators	The model compared TMS with sham TMS and pharmacotherapy as usual
S6: Model type	Cost-utility analysis
S7: Time horizon	One-year time horizon chosen to capture disease progression. Acute and taper phase outcomes were captured at 6 weeks using decision tree and the Markov was extended for 1 year
S8: Disease states/pathways	The pathways/health states modelled are relevant
S9: Cycle length	Three-monthly cycles and no justification was given for the chosen cycle length

Dimension of quality	Comments
Data	
D1: Data identification	Data were systematically sourced, clearly described and justified by the authors
D2: Pre-model data analysis	The authors did not report any pre-model analysis. Data derived from different meta-analysis for different parameters was directly implemented in the model
D2a: Baseline data	Events reported in three different studies were used to populate the model
D2b: Treatment effects	Treatment effect reported in three different studies (referenced) including the STAR*D were used to populate the model
D2c: Costs	Adverse event costs were excluded, otherwise included costs were appropriate and resources clearly stated
D2d: Quality-of-life weights (utilities)	Derived from literature and clearly referenced
D3: Data incorporation	The authors clearly described how data were used in the model
D4: Assessment of uncertainty	The assessment of sensitivity was thorough and robust as both deterministic and PSA were undertaken
D4a: Methodological	The authors used appropriate analytical methods for the decision problem
D4b: Structural	The authors described deterministic sensitivity analysis and PSA was also undertaken
D4c: Heterogeneity	Heterogeneity was assessed by reporting results separately by depression severity, i.e. mild, moderate and severely depressed
D4d: Parameter	OWSA and PSA were undertaken
Consistency	
C1: Internal consistency	The model seems to be mathematically sound, with no obvious inconsistencies
C2: External consistency	The results of the model are applicable to patients with MDD. Resource-use data were obtained from individual patient surveys. Overall, the study findings and conclusions cannot be directly generalised to the UK setting but nonetheless the methodology was informative to our economic model

Appendix 9 Table of excluded health economics studies with rationale

Cost-effectiveness studies

Bibliographic information	Reason for exclusion
Bosmans J, de Bruijne M, van Hout H, van Marwijk H, Beekman A, Bouter L, <i>et al.</i> Cost-effectiveness of a disease management program for major depression in elderly primary care patients. <i>J Gen Int Med</i> 2006; 21 :1020–6	Patient population: First line of therapy. Also the intervention/comparators, i.e. GP trained to screen and treat vs usual care
Wells KB, Schoenbaum M, Duan N, Miranda J, Tang L, Sherbourne C. Cost-effectiveness of quality improvement programs for patients with subthreshold depression or depressive disorder. <i>Psychiatr Serv</i> 2007; 58 :1269–78	Patient population: Subthreshold depression and dysthymic disorder, interventions are quality improvements programmes
Perlis RHP. When is pharmacogenetic testing for antidepressant response ready for the clinic? A cost-effectiveness analysis based on data from the STAR*D study. <i>Neuropsychopharmacology</i> 2009; 34 :2227–36	Patient population: First-line MDD, the study assessed the cost-effectiveness of pharmacogenetic testing
Armstrong EP, Skrepnek GH, Erder MH. Cost-utility comparison of escitalopram and sertraline in the treatment of major depressive disorder. <i>Curr Med Res Opin</i> 2007; 23 :251–8	Patient population: First line of therapy (although does include data on the probability of augmentation)
Sicras-Mainar A, Navarro-Artieda R, Blanca-Tamayo M, Gimeno-delaFuente V, Salvatella-Pasant J. Comparison of escitalopram vs. citalopram and venlafaxine in the treatment of major depression in Spain: clinical and economic consequences. <i>Curr Med Res Opin</i> 2010; 26 :2757–64	Patient population: First line of therapy
Einarson TR, Addis A, Iskedjian M. Pharmacoeconomic analysis of venlafaxine in the treatment of major depressive disorder. <i>Pharmacoeconomics</i> 1997; 12 :286–96	Patient population: First line of therapy
van Baardewijk M, Vis PM, Einarson TR. Cost effectiveness of duloxetine compared with venlafaxine-XR in the treatment of major depressive disorder. <i>Curr Med Res Opin</i> 2005; 21 :1271–9	Patient population: Not treatment resistant were excluded, also interventions
Casciano J, Arikian S, Tarride J, Doyle JJ, Casciano R. A pharmacoeconomic evaluation of major depressive disorder. <i>Epidemiol Psychiatr Soc</i> 1999; 8 :220–31	Patient population: First line of therapy
Trivedi MH, Wan GJ, Mallick R, Chen JL, Casciano R, Geissler EC, <i>et al.</i> Cost and effectiveness of venlafaxine extended-release and selective serotonin reuptake inhibitors in the acute phase of outpatient treatment for major depressive disorder. <i>J Clin Psychopharmacol</i> 2004; 24 :497–506	Patient population: Not treatment resistant. There is no augmentation
Wade AG, Fernandez JL, Francois C, Hansen K, Danchenko N, Despiegel N. Escitalopram and duloxetine in major depressive disorder: a pharmacoeconomic comparison using UK cost data. <i>Pharmacoeconomics</i> 2008; 26 :969–81	Patient population: First line of therapy
Freeman H, Arikian S, Lenox-Smith A. Pharmacoeconomic analysis of antidepressants for major depressive disorder in the United Kingdom. <i>Pharmacoeconomics</i> 2000; 18 :143–8	Patient population: Not treatment resistant; also first line then titration and/or augmentation upon failure of first line

Bibliographic information	Reason for exclusion
Revicki DA, Siddique J, Frank L, Chung JY, Green BL, Krupnick J, <i>et al.</i> Cost-effectiveness of evidence-based pharmacotherapy or cognitive behavior therapy compared with community referral for major depression in predominantly low-income minority women. <i>Arch Gen Psychiatry</i> 2005; 62 :868–75	Patient population: First line of therapy
Lenox-Smith A, Conway P, Knight C. Cost effectiveness of representatives of three classes of antidepressants used in major depression in the UK. <i>Pharmacoeconomics</i> 2004; 22 :311–19	Patient population: First line of therapy
Lenox-Smith A, Greenstreet L, Burslem K, Knight C. Cost effectiveness of venlafaxine compared with generic fluoxetine or generic amitriptyline in major depressive disorder in the UK. <i>Clin Drug Invest</i> 2009; 29 :173–84	Patient population: First line of therapy
Armstrong EP, Malone DC, Erder MH. A Markov cost-utility analysis of escitalopram and duloxetine for the treatment of major depressive disorder. <i>Curr Med Res Opin</i> 2008; 24 :1115–21	Patient population: First-line MDD
Revicki DA, Brown RE, Keller MB, Gonzales J, Culpepper L, Hales RE, <i>et al.</i> Cost-effectiveness of newer antidepressants compared with tricyclic antidepressants in managed care settings. <i>J Clin Psychiatry</i> 1997; 58 :47–58	Patient population: Not treatment resistant at baseline. Only 12% are assumed to have TRD in the model
Machado M, Lopera MM, az-Rojas J, Jaramillo LE, Einarson TR, Universidad Nacional de Colombia Pharmacoeconomics Group. Pharmacoeconomics of antidepressants in moderate-to-severe depressive disorder in Colombia. <i>Pan American Journal of Public Health</i> 2008; 24 :233–9	Patient population: First episode of moderate to severe depression also first line then titration and/or augmentation upon failure of first line
Lave JR, Frank RG, Schulberg HC, Kamlet MS. Cost-effectiveness of treatments for major depression in primary care practice. <i>Arch Gen Psychiatry</i> 1998; 55 :645–51	Patient population: First line of therapy
Gibson TBJ. Cost burden of treatment resistance in patients with depression. Cost-utility comparison of escitalopram and sertraline in the treatment of major depressive disorder. <i>Am J Manag Care</i> 2010; 16 :370–7	Not a cost-effectiveness study (a modelling study showing that patients with TRD have higher medical costs than non-TRD patients)

Excluded quality-of-life studies

Bibliographic information	Reason for exclusion
Cipriani A, Smith K, Burgess S, Carney S, Goodwin G, Geddes J, <i>et al.</i> Lithium vs. antidepressants in the long-term treatment of unipolar affective disorder. <i>Cochrane Database Syst Rev</i> 2006; 4 :CD003492	Review
Armstrong EPM. A Markov cost-utility analysis of escitalopram and duloxetine for the treatment of major depressive disorder. <i>Curr Med Res Opin</i> 2008; 24 :1115–21	CEA study, which used QoL data from Revicki which was included
Benedict A, Arellano J, De Cock E, Baird JE-MA, Benedict AAB. Economic evaluation of duloxetine vs. serotonin selective reuptake inhibitors and venlafaxine XR in treating major depressive disorder in Scotland. <i>J Affect Disord</i> 2010; 120 :1–3	Included in the CEA review
Revicki AD, Hanlon J, Martin S, Gyulai L, Ghaemi NS, Lynch F, <i>et al.</i> Patient-based utilities for bipolar disorder-related health states. <i>J Affective Dis</i> 1998; 87 :203–10	Conducted in bipolar patients, a similar one in unipolar was included
Leelahanj T, Leelahanj T. The cost-effectiveness of aripiprazole as adjunctive therapy in major depressive disorder: Thai economic model. <i>J Med Assoc Thai</i> 2010; 93 (Suppl. 6):43–50	Included in the CEA review
Ng FD. Combination pharmacotherapy in unipolar depression. <i>Expert Rev</i> 2006; 6 :1049–60	A review on pharmacotherapy and has no QoL data
Gismondi RL. Effect of adjunctive aripiprazole on quality of life in patients with major depressive disorder: pooled data from three clinical trials. European Psychiatry 2010, conference (various pages)	Conference abstract
Keitner GIG. A randomized, placebo-controlled trial of risperidone augmentation for patients with difficult-to-treat unipolar, non-psychotic major depression. <i>J Psychiatr Res</i> 2009; 43 :205–14	Conference abstract
Franco Martn M, Figueira ML. Quetiapine XR monotherapy and quetiapine XR+ongoing antidepressant vs. lithium+ongoing AD for Stage II treatment-resistant major depressive disorder. <i>Eur Neuropsychopharmacol</i> 2010; 20 :347	Conference abstract
King M, Sibbald B, Ward E, Bower P, Lloyd M, Gabbay M, <i>et al.</i> Randomised controlled trial of nondirective counselling, cognitive-behaviour therapy and usual general practitioner care in the management of depression as well as mixed anxiety and depression in primary care. <i>Health Technol Assess</i> 2000; 4 (19)	CEA study, which reports treatment-specific utilities
Peveler R, Kendrick T, Buxton M, Longworth L, Baldwin D, Moore M, <i>et al.</i> A randomised controlled trial to compare the cost-effectiveness of tricyclic antidepressants, selective serotonin reuptake inhibitors and lofepramine. <i>Health Technol Assess</i> 2005; 9 (16)	CEA study, which reports treatment-specific utilities
Lenert LA. The reliability and internal consistency of an internet capable program for measuring utilities. <i>Qual Life Res</i> 2000; 9 :811–17	A feasibility study testing the reliability of a software

CEA, cost-effectiveness analysis.

A decorative graphic consisting of numerous thin, parallel green lines that curve from the left side of the page towards the right, creating a sense of movement and depth.

**EME
HS&DR
HTA
PGfAR
PHR**

Part of the NIHR Journals Library
www.journalslibrary.nihr.ac.uk

This report presents independent research funded by the National Institute for Health Research (NIHR). The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health

Published by the NIHR Journals Library