Comparative efficacy, acceptability and cost effectiveness of pharmacological treatments for depression in adults with bipolar disorder: a network meta-analysis and modelling economic study [study protocol]


1 Research Department of Clinical, Educational & Health Psychology, University College London, London, UK
2 Department of Psychiatry, University of Oxford, Oxford, UK
3 School of Social and Community Medicine, University of Bristol, Bristol, UK
4 Institute of Mental Health, University of Nottingham, Nottingham, UK
5 National Guideline Centre, Royal College of Obstetricians and Gynaecologists, London, UK
6 Center for Clinical Trials and Evidence Synthesis, Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, 21205, USA
7 National Collaborating Centre for Mental Health, Royal College of Psychiatrists, London, UK
BACKGROUND

Bipolar disorder is a cyclical mood disorder that involves periods of depression or elevated mood (mania or hypomania), alternated with periods of full recovery or much improved function. Although the distinguishing feature of bipolar disorder is the experience of hypomania or mania, on average, people with bipolar disorder experience more depressive than manic episodes and spend more time with syndromal or subsyndromal depressive episodes than with mania (Judd et al., 2002 & 2003), with depressive symptoms having more profound effects on social functioning compared with manic symptoms (Morris et al., 2013). Together with psychological therapies, pharmacological interventions are the mainstay treatment to manage bipolar depression in a clinical setting.

In 2014, the National Institute for Health and Care Excellence (NICE) published updated guidance on the management of bipolar disorder in England (NICE, 2014). The guideline included a network meta-analysis (NMA) of pharmacological treatments for depression, which considered 27 double-blind randomised controlled trials (RCTs) and 9006 participants. The NMA subsequently informed a modelling economic study assessing the cost effectiveness of interventions available in the UK. Both the NMA and the economic analysis identified valproate as the optimal (most effective and cost-effective) option for the treatment of depression. Nevertheless, the guideline recommended use of combined olanzapine + fluoxetine or quetiapine for the management of depression, which were the second and third most effective and cost-effective options available in the UK, respectively, as evidence on valproate was very limited (valproate had been tested in N=48 in total).

Another published NMA, including 29 double-blind RCTs and 8331 participants, recommended combined olanzapine + fluoxetine as first-line treatment for bipolar depression based on the balance of likelihood of response and lower likelihood of withdrawal, and also found that olanzapine or quetiapine could be considered alternative first-line treatments (Taylor et al., 2014). In this NMA, antidepressants were assessed in classes [selective serotonin re-uptake inhibitors (SSRIs), monoamine oxidase inhibitors (MAOIs), tricyclic
antidepressants (TCAs)], whereas mood stabilisers and antipsychotics were examined independently, which may have implications for decision-making. More recently, guidelines from the British Association for Psychopharmacology and the Royal Australian and New Zealand College of Psychiatrists recommended use of quetiapine, lurasidone or olanzapine as first-line treatment of depression in patients not already taking long-term treatment for bipolar disorder (Goodwin et al., 2016; Malhi et al., 2015).

People with bipolar disorder and their families incur considerable costs associated with provision of health and social care as well as productivity losses (due to time-off work, unemployment or suicides), which were estimated at £5.2 billion in the UK in 2007, expected to rise up to £8.2 billion in 2026 (McCrone et al., 2008), €5.8 billion in Germany in 2002 (Runge & Grunze, 2004), $1.8 billion in the Netherlands in 2002 (Hakkaart-van Roijen et al., 2004), and $151 billion in the US in 2009 (Dilsaver, 2011). The recommendation as first line treatment of new drugs (such as lurasidone) that have been just licensed by regulatory agencies can have important consequences not only in terms of clinical effectiveness, but also in terms of economic implications. Efficient use of available health and social care resources must therefore be ensured, to maximise the health benefits for people with bipolar disorder and, at the same time, reduce the global burden to society.

**OBJECTIVES**

To compare the efficacy, acceptability and cost-effectiveness of pharmacological interventions in the treatment of acute depressive episodes in adults with bipolar disorder, by updating the NMA and the economic analysis that were undertaken for the NICE guideline (NICE, 2014).

**METHODS**

**CLINICAL ANALYSIS**

Types of studies
Double-blind RCTs comparing different pharmacological treatments (either as monotherapy or combined with another drug) with placebo or another active pharmacological intervention for the acute treatment of bipolar depression will be included. Quasi-RCTs, such as trials in which allocation is determined by alternation or date of birth and single-blind studies will be excluded.

**Types of participants**

Participants aged 18 or older, of both sexes, with a primary diagnosis of bipolar I disorder or bipolar II disorder, who are experiencing an acute episode of depression, diagnosed according to any of the following operationalized criteria: Research Diagnostic Criteria, DSM-III, DSM-III-R, DSM-IV, DSM-IV-TR, DSM-IV or ICD-10. Operationalized criteria essentially resembling these official ones will also be eligible. Studies conducted in primary, secondary, and tertiary health care settings and social care settings will be considered. Studies recruiting specifically patients with medical or psychiatric co-morbidities will be excluded.

**Types of interventions**

We will consider pharmacological interventions that are currently indicated or used in the treatment of depression in people with bipolar disorder. We will exclude studies or arms if participants are receiving non-pharmacological treatments and over-the-counter drugs. We will also exclude drugs that are tested in the treatment of bipolar disorder but have not been manufactured primarily for the management of mental health disorders or other disorders of the central nervous system (such as, non-steroidal anti-inflammatory drugs or antibiotic drugs).

We will include pharmacological interventions alone or in combination, administered in fixed or flexible doses within the therapeutic range recommended by the British National Formulary.
(BNF)¹ or by the Food and Drug Administration (FDA)² if relevant information is not available in the BNF. RCTs evaluating classes of pharmacological interventions (e.g. SSRIs, TCAs, atypical antipsychotics, mood stabilisers), rather than independent drugs, will not be considered. We will search for RCTs that evaluate one or more of the following pharmacological interventions: antidepressants (agomelatine, amitriptyline, brofaromine, bupropion, citalopram, clomipramine, desvenlafaxine, duloxetine, escitalopram, fluoxetine, fluvoxamine, imipramine, maprotiline, mianserin, mirtazapine, moclobemide, nefazodone, paroxetine, phenelzine, reboxetine, sertraline, tianeptine, tranylcypromine, trazodone, venlafaxine, vortioxetine), antipsychotics (aripiprazole, asenapine, brexipiprazole, cariprazine, chlorpromazine, clozapine, haloperidol, iloperidone, lurasidone, olanzapine, paliperidone, quetiapine, risperidone, ziprasidone, zotepine), lithium, anti-epileptic drugs (carbamazepine, gabapentin, lamotrigine, tiagabine, topiramate, valproate and valproic acid), and other drugs with different mechanisms of action on the central nervous system (for example pramipexole, etc).

We will obtain information about the interventions of interest either from head-to-head or placebo-controlled trials. We assume that any patient that meets all inclusion criteria is, in principle, equally likely to be randomized to any of the interventions in the synthesis comparator set.

We will not consider in the NMA interventions that are not listed above, unless they act as the sole connectors of the interventions of interest (or their combinations) in the network. In this case, interventions not listed above will be included in the NMA but will not form part of the decision problem (Dias et al., 2013a).

**Types of outcome measures**

² [http://www.fda.gov/Drugs/default.htm](http://www.fda.gov/Drugs/default.htm)
(1) Efficacy: the proportion of patients who demonstrated at least 50% reduction in depressive symptoms at the end of acute treatment, as measured by the Hamilton Depression Rating Scale (HAMD) (Hamilton, 1960), the Montgomery Åsberg Depression Rating Scale (MADRS) (Montgomery and Asberg, 1979) or the Quick Inventory of Depressive Symptomatology (QIDS) (Rush et al., 2003).

(2) Acceptability: the proportion of patients who left the study early, before completion of acute treatment, due to any cause.

**Time horizon**

We will include trials with duration of acute treatment of 6 weeks minimum and 12 weeks maximum. We will include outcomes at study endpoint.

**Network connections**

A network diagram for all outcomes will be constructed to explore whether all interventions are connected to the network. If more than one networks are formed, then separate NMAs will be conducted for each network, as long as the network contains at least 3 interventions that are part of the decision problem.

The network developed for the NICE guideline on bipolar disorder, which will be updated in this NMA update, is shown in Figure 1.
Search methods for identification of studies

The search strategy will be designed to detect all published and unpublished RTCs comparing treatments against each other or with placebo.

1. Electronic searches

We will search MEDLINE, PreMEDLINE, PsycINFO, Embase, the Cochrane Central Register of Controlled Trials (CENTRAL) and the Cumulative Index to Nursing and Allied Health Literature (CINAHL). No language restriction will be applied to the search.
We will update searches conducted for the Bipolar disorder guideline developed by the National Institute for Health and Care Excellence (NICE, 2014) which searched for literature from databases’ inception to January 2014. We will search the databases from January 2014 to 31 January 2017 using terms for the population combined with a highly sensitive filter for RCTs. We will use the following search strategy for Medline which we will adapt for use across the other databases.

**Medline search strategy- OVID platform**

```
1 exp "bipolar and related disorders"/ or cyclothymic disorder/
2 (((bi polar or bipolar) adj5 disorder*) or ((cyclothymi* or rapid or ultradian) adj5 cycl*)
   or hypomani* or mania* or manic* or mixed episode* or rcbd).ti,ab.
3 or/1-2
4 depression/ or exp depressive disorder/
   (depress$ or dysphori$ or dysthymi$ or melancholi$ or seasonal affective disorder$).ti,ab.
5 or/4-5
6 3 and 6
7 (bipolar adj5 depression).ti,ab.
8 7 or 8
9 exp clinical trial/ or exp "clinical trials as topic"/ or cross-over studies/ or double-blind
   method/ or placebos/ or random allocation/ or single-blind method/
10 (clinical adj2 trial$).ti,ab.
11 (crossover or cross over).ti,ab.
12 (((single$ or doubl$ or trebl$ or tripl$) adj2 blind$) or mask$ or dummy or
   doubleblind$ or singleblind$ or trebleblind$ or tripleblind$).ti,ab.
13 (placebo$ or random$).ti,ab.
14 animal$/ not human$/
15 or/10-14 not 15
16 or/10-14 not 15
17 9 and 16
```

Other search methods will involve:

(a) scanning the reference lists of all eligible publications for published reports and citations of unpublished research;
(b) searching the NICE guideline on Bipolar Disorder for evidence on pharmacological interventions for adults with bipolar depression that was included in the respective systematic review of the guideline; we will look at

i) the studies included in the guideline NMA to confirm that these meet our inclusion criteria and to explore whether these report additional outcome data that are useful in our proposed NMA, which had not been extracted for the guideline NMA (e.g. we propose to extract QIDS data that were not considered in the guideline NMA)

ii) the list of excluded studies from the guideline NMA to determine whether any of these meet the inclusion criteria for this systematic review

(c) searching clinicaltrials.gov and WHO-ICTRP for unpublished studies for all years;

(d) contacting the manufacturers of newer drugs, including brexpiprazole, cariprazine, lurasidone and vortioxetine, to check that all relevant studies, either published or unpublished, have been considered for inclusion

Study selection will follow the PRISMA guideline reporting flow for information in systematic reviews of the literature that incorporate network meta-analyses (Hutton et al, 2015).

Data collection and analysis

Selection of studies

Two review authors will examine the titles/abstracts of all publications obtained through the search strategy. Full reports of all the studies identified by either of the review authors will then be obtained and inspected by the same two review authors to identify trials meeting the inclusion criteria. Conflicts of opinion regarding eligibility of a study will be discussed with a third review author, having retrieved the full report(s) and consulted the authors if necessary, until consensus is reached. External subject or methodological experts will be consulted if necessary.

Assessment of risk of bias in included studies
Risk of bias will be assessed for each included study using the Cochrane Collaboration 'risk of bias' tool as a model (Higgins & Green, 2011). The following 6 domains will be considered:

1. Sequence generation: was the allocation sequence adequately generated?
2. Allocation concealment: was allocation adequately concealed?
3. Blinding of participants, personnel and outcome assessors for each main outcome: was knowledge of the allocated treatment adequately prevented during the study?
4. Incomplete outcome data for the primary outcomes: were incomplete outcome data adequately addressed?
5. Selective outcome reporting: are reports of the study free of suggestion of selective outcome reporting?
6. Other bias.

A description of what was reported to have happened in each study will be provided, and a judgment on the risk of bias will be made for each domain, based on the following three categories:

- High risk of bias
- Low risk of bias
- Unclear risk of bias.

Two independent review authors will assess the risk of bias in selected studies. Any disagreement will be resolved through discussion and in consultation with the principal investigators. Where necessary, the authors of the studies will be contacted for further information.

**Data extraction and synthesis**

We will generate descriptive statistics for trial and study population characteristics across all eligible trials, describing inclusion and exclusion criteria, the types of comparisons (including dose, frequency and duration of interventions and comparators) and some important variables, either clinical or methodological (such as diagnosis [percentage of patients with
bipolar I disorder], year of publication, country, gender [percentage of female], age [mean], severity of illness, sponsorship, clinical setting).

The following data will be extracted from each study for each outcome of interest:

**Efficacy:**

- Number of people who have responded by the end of acute treatment, with response being defined as 50% reduction in depressive symptoms, measured on any of the HAMD, MADRS or QIDS. All available binary response data on any of these scales will be collected, so if a study reports response data on more than one of these scales, all these data will be extracted.

- Baseline and endpoint or change score data (mean, standard deviation [SD]) on the HAMD, MADRS and QIDS scales; these will be extracted only from studies that do not provide binary response data, in order to inform efficacy. All available continuous data on any of these scales will be collected, so if a study provides continuous data on more than one of these scales, all these data will be extracted. We will extract continuous data on randomised patients (intention-to-treat [ITT] data, reported following imputation of missing data) and/or continuous data on study completers (completers’ data), depending on availability of continuous data in each study; where both ITT and completers’ data are available, both types of data will be extracted.

**Acceptability:**

- Number of participants leaving the study early, before completion of acute treatment, due to any cause

**Network meta-analysis (NMA):**

Random-effects NMA taking into account the heterogeneity of treatment effects across studies will be conducted in a Bayesian framework using Markov Chain Monte Carlo simulation methods implemented in WinBUGS 1.4.3 (Lunn et al., 2000; Spiegelhalter, 2001). Data on
each endpoint outcome will be combined on the log-odds ratio scale and non-informative priors will be assigned. To test whether prior estimates have an impact on the results, three chains with different initial values will be run simultaneously. Convergence will be assessed by inspection of the Brooks-Gelman–Rubin diagnostic plot (Brooks & Gelman, 1998), and by inspecting the trace plots. Goodness of fit will be tested using the posterior mean of the residual deviance, which will be compared with the number of data points in the model (Dias et al., 2011). Inconsistency, i.e. differences between the direct and indirect evidence on comparisons for which both are available, will be assessed globally by comparing the fit and estimates of between-study heterogeneity of the model not assuming consistency to the NMA model (Dias et al., 2013b). Depending on the structure of the network, further checks will be conducted using a node-split approach implemented in R using the gemtc package (Dias et al., 2010).

Results for the comparative efficacy, acceptability and tolerability will be presented by odds ratio estimates and 95% confidence intervals. We will also evaluate the ranking of each treatment and 95% credible intervals for efficacy and tolerability and the posterior probability of which intervention is the best (in terms of efficacy and tolerability) regimen, the second best, the third best, and so on. For the economic analysis, posterior absolute probabilities of discontinuation and response given no discontinuation for each intervention at end of treatment will be estimated.

Standard models for NMA with binary outcomes will be used for three outcomes: (a) discontinuation due to any reason (drop-out rate); (b) response in all randomised patients and (c) response in randomised patients who did not discontinue treatment (completers’ response). Outcomes (a) and (b) will be used to inform the clinical analysis. Outcomes (a) and (c) will be used to inform the economic analysis.

*Synthesis of efficacy data extracted from each study arm:*
The log-odds ratio of response for each study will be calculated as follows:

(1) For studies reporting the number of responders on only one of the HAMD, MADRS or QIDS scales, those data will be used in the analysis. We will assume that patients who dropped out did not respond to treatment.

(2) For studies reporting the number of responders on more than one of the HAMD, MADRS and QIDS, the log-odds ratio of completers’ response given by each scale will be averaged and the standard error of the log-odds ratio will be calculated as the average of the standard errors on each scale.

(3) For studies not reporting the number of responders but reporting the mean and SD on one of the scales (HAMD or MADRS or QIDS), the within-study standardised mean difference (SMD) and its variance will be calculated according to the Hedges’ g and other standard formulae (Higgins & Green, 2011), transformed to a notional log-odds ratio for response using the following formula:

\[ \text{LOR} = \frac{\pi}{\sqrt{3}} \times \text{SMD} \] (Chinn, 2000).

(4) For studies not reporting the number of responders but reporting the mean and SD on more than one of the HAMD, MADRS and QIDS scales, the within-study SMD on each scale and their standard errors will be calculated as above, and then averaged. This combined SMD and its variance (the standard error squared) will be transformed to a notional log-odds ratio for response as above.

Regarding (3) and (4), in studies where when both ITT and completers’ continuous data are reported, we will use ITT continuous data to estimate ITT response, and completers’ continuous data to estimate completers’ response. If only one type of continuous data (i.e. either ITT or completers’) is reported in a study, we will consider the available data and check the sensitivity of the results to the exclusion of studies reporting only continuous data, as described below.
To obtain absolute probabilities of discontinuation and response given no discontinuation that are required for the economic analysis, it is necessary to obtain a baseline treatment effect for placebo to which the relative treatment effects are applied. The baseline (placebo) probabilities of discontinuation and response given no discontinuation will be estimated by performing a meta-analysis of selected placebo arms from placebo-controlled RCTs included in the NMA for the two outcomes of interest: discontinuation and response given not discontinued. Selection of placebo arms will be made on the basis of the trials’ representativeness of the study population in the UK in terms of baseline rates of discontinuation and response, as judged by the research team. Subsequently, the predictive distributions of the log-odds of discontinuation and on response given not discontinued on placebo in a future trial will be estimated and form the baseline treatment effect for placebo, to which the relative treatment effects will be applied.

**Sensitivity and subgroup analyses**

Sensitivity analyses will be conducted to address risk of bias relating to

(a) assessment of each study according to the Cochrane Risk of Bias Tool: studies assessed at high risk of bias for blinding and/or allocation concealment will be removed from the NMA and results will be compared with the main analysis

(b) exclusion of studies reporting only continuous data on HAMD, MADRS and/or QIDS: studies included in the NMA that only report continuous data on HAMD, MDRS and/or QIDS will be removed from the NMA and efficacy results of this analysis (which will include exclusively binary response data) will be compared with those of the main analysis (which will include binary response data and continuous data from studies not reporting binary response data).

Subgroup analyses will be conducted by bipolar disorder subtype (I or II) as the efficacy and acceptability to pharmacological treatment may differ between the two subtypes of the disorder.
ECONOMIC ANALYSIS

We will update the economic model that was undertaken for the NICE guideline on Bipolar Disorder (NICE, 2014) to assess the cost effectiveness of pharmacological interventions for the management of acute depressive episodes in adults with bipolar disorder from the perspective of the National Health Service (NHS) and Personal Social Services (PSS) in the UK. The model was a decision-tree developed in Microsoft Excel 2010. The model update will follow the structure of the NICE guideline model and will consider the events of treatment discontinuation and moving to a second treatment option or no treatment, full or partial response to treatment, and the risk of manic or depressive relapse. The time horizon of the model will be the sum of duration of acute treatment, determined by the mean duration of trials included in the updated NMA, plus 12 weeks of continuation of drug treatment, prior to initiation of long-term pharmacological maintenance treatment, as in the NICE guideline model (NICE, 2014).

The economic model will consider only drugs licensed in the UK, alone or in combination, which will have been found to be ranking higher than placebo in terms of efficacy. No pharmacological treatment (reflected in treatment with placebo) will also be included as a treatment option, in order to assess the cost effectiveness of active interventions versus a non-specific medical management (used as a benchmark).

The main clinical input parameters for the model will be obtained from the NMA. Other event probabilities in the model will be obtained from the model developed for the NICE guideline on Bipolar Disorder (NICE, 2014).

The measure of outcome of the economic analysis will be the Quality Adjusted Life Year (QALY). Utility data required for the estimation of QALYs will be obtained from the systematic review of relevant studies undertaken for the NICE guideline on Bipolar Disorder.
(NICE, 2014), following NICE recommendations on the selection of utility values for use in cost-utility analysis (NICE, 2013).

Cost elements will include drug acquisition costs, contacts with healthcare professionals’ time, laboratory testing required as part of pharmacological interventions, hospitalisation or management by crisis resolution and home treatment teams (CRHTTs) for a proportion of people not responding to treatment. Costs will be estimated by combining relevant resource use for each arm and pathway of the model with appropriate unit costs. Resource use will be estimated based on the resource use reported in RCTs included in the NMA, combined with estimates reported in the NICE guideline on Bipolar Disorder that were used in the guideline economic model (NICE, 2014). Unit costs will be sought from national sources. Regarding drug acquisition costs, the lowest reported national prices will be selected and used in the analysis; where available, costs of generic forms will be considered. All costs will be uplifted to 2016 GBP prices.

Deterministic and probabilistic analyses will be conducted. Model input parameters will be assigned appropriate probabilistic distributions to reflect the uncertainty characterising the available clinical and cost data. One-way deterministic sensitivity analyses will be carried out to explore the robustness of the results under different input estimates and assumptions in the model.

Incremental analysis will be conducted and results will be expressed as Incremental Cost Effectiveness Ratios, expressing the difference in total costs between two interventions and divided by the difference in their effectiveness (QALYs). In addition to ICERs, the mean net monetary benefit (NMB) of each intervention will be presented (Fenwick et al., 2001). Results will be summarised in the form of cost effectiveness planes and cost effectiveness acceptability curves (Fenwick et al., 2001).
The update of the economic model will be undertaken by one health economist and checked for errors and internal consistency by a second modeller. The model will also be tested for logical consistency by setting input parameters to null and extreme values and examining whether results change in the expected direction. The results will be discussed within the review team for their plausibility.
REFERENCES


recommendations from the British Association for Psychopharmacology. Journal of Psychopharmacology, DOI: 10.1177/0269881116636545


