

# **A systematic review and critical appraisal of economic evaluations of pharmacological interventions for people with bipolar disorder**

**Running title:** systematic review of cost effectiveness of pharmacotherapy in bipolar disorder

**Authors:** Ifigeneia Mavranouzouli<sup>1</sup>, Joran Lokkerbol<sup>2,3</sup>

1 National Guideline Alliance (NGA), Centre for Outcomes Research and Effectiveness, Research Department of Clinical, Educational & Health Psychology, University College London, London, UK

2 Centre of Economic Evaluation, Trimbos Institute (The Netherlands Institute of Mental Health and Addiction), Utrecht, The Netherlands

3 Rob Giel Research Centre, University Medical Centre Groningen, Groningen, The Netherlands.

## **Corresponding author:**

Ifigeneia Mavranouzouli

National Guideline Alliance (NGA)

Centre for Outcomes Research and Effectiveness (CORE)

Research Department of Clinical, Educational & Health Psychology

1-19 Torrington Place

London WC1E 7HB, UK

e-mail: [i.mavranouzouli@ucl.ac.uk](mailto:i.mavranouzouli@ucl.ac.uk)

**Conflicts of interest:** none

**Funding:** none

**Wordcount:** 8,142 excluding abstract, tables and references

## **Abstract**

**Background:** Bipolar disorder is a chronic mood disorder that causes substantial psychological and financial burden. Various pharmacological treatments are effective in the management and the prevention of acute episodes of bipolar disorder. In an era of tighter healthcare budgets and a need for more efficient use of resources, several economic evaluations have evaluated the cost effectiveness of treatments for bipolar disorder. The objective of this study was to systematically review and appraise published economic evaluations of pharmacological interventions for bipolar disorder.

**Methods:** A systematic search combining search terms specific to bipolar disorder with a health economics search filter was conducted on 6 bibliographic databases (EMBASE, MEDLINE, PsycINFO, HTA, NHS EED, CENTRAL) between 1 January 1990 and 18 December 2015 in order to identify trial- or model-based full economic evaluations of pharmacological treatments of any phase of the disorder. Studies that met inclusion criteria were critically appraised using the Quality of Health Economics Studies (QHES) checklist and synthesised in a narrative way.

**Results:** The review included 19 economic studies. The studies varied regarding the type and number of interventions assessed, the study design, the phase of treatment (acute or maintenance), the source of efficacy data and the method for evidence synthesis, the outcome measures, the time horizon and the countries/settings in which the studies were conducted. The study quality was variable, but the majority of studies were of high or fair quality.

**Conclusion:** Pharmacological interventions are cost-effective compared with no treatment in the management of bipolar disorder, both in the acute and in maintenance phase; however it is difficult to draw safe conclusions on the relative

cost effectiveness between drugs, due to differences across studies and limitations characterising many of them. Future economic evaluations need to consider the whole range of treatment options available for the management of bipolar disorder and adopt appropriate methods for evidence synthesis and economic modelling, to explore more robustly the relative cost effectiveness of pharmacological interventions for people with bipolar disorder.

### **Key points for decision-makers**

- Existing evidence, coming primarily from model-based economic evaluations, suggests that pharmacological interventions are cost-effective compared with no treatment in the management of people with bipolar disorder
- The relative cost effectiveness between drugs is difficult to be established due to differences in the interventions examined, the overall methods employed, the data sources and the settings across economic studies, as well as methodological caveats characterising many of them
- Model-based economic evaluations in the area of bipolar disorder need ideally to include all available treatment options and follow robust methods for evidence synthesis. Models should have appropriate structure, consider key events in the course of bipolar disorder and have adequate time horizons, ideally over lifetime, in order to capture the long-term benefits, harms and costs associated with evaluated interventions.

Bipolar disorder (BD) is a cyclic mood disorder that involves periods of low (depression) and high (mania or hypomania) mood, energy and activity levels, alternated with periods of full recovery or much improved function (euthymia) [1]. It is further distinguished in BD type I, in which the person experiences full-blown manic episodes (most commonly interspersed with episodes of major depression), and BD type II, in which the person has depressive episodes and less severe manic (hypomanic) episodes [2]. Mixed episodes, where the symptoms of depression and mania or hypomania occur at the same time to a marked degree, or rapid cycling, where people experience at least 4 syndromal depressive, manic, hypomanic or mixed episodes within a 12-month period, are also expressions of the disorder [2].

BD is one of the leading causes of disability due to mental and behavioural disorders worldwide, following unipolar depression, anxiety, substance misuse, alcohol use and schizophrenia. According to the World Health Organization (WHO) Global Burden of Disease study, from 1990 to 2010 there was a 40.9% increase in disability-adjusted life years (DALYs) attributable to bipolar disorder over the globe [3]. People with bipolar disorder incur considerable costs associated with provision of health and social care as well as productivity losses (due to time-off work, unemployment or suicide), which were estimated at £5.2 billion in the UK in 2007, expected to rise up to £8.2 billion in 2026 [4], €5.8 billion in Germany in 2002 [5], \$1.8 billion in the Netherlands in 2002 [6], and \$151 billion in the US in 2009 [7].

A range of pharmacological and psychological interventions are effective in the management and the prevention of acute episodes of BD. Drugs such as lithium, antipsychotics (e.g. aripiprazole, asenapine, haloperidol, lurasidone, olanzapine, quetiapine, risperidone), antidepressants (e.g. fluoxetine, imipramine, paroxetine, venlafaxine) and antiepileptics (such as carbamazepine, lamotrigine and valproate, either as valproic acid or sodium valproate) are being used, alone or in combination, for this purpose [8]. In an era of tight healthcare budgets, where efficient use of limited resources is imperative, there is a large number of economic studies that have evaluated the cost effectiveness of drugs used in the treatment of BD.

The objective of this study was to systematically review and appraise published economic evaluations of pharmacological interventions used in the management of BD, in order to identify cost-effective interventions and factors that have a significant impact on cost effectiveness in this disease area. Moreover, the study aimed to explore methodological strengths and limitations of existing economic literature, and recommend good practice points when conducting and appraising economic studies in the area of BD.

## **1. Methods**

This systematic review was conducted according to the PRISMA recommendations for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions [9].

### **1.1 Search strategy**

A systematic search of the literature was conducted in standard bibliographic databases (EMBASE, MEDLINE and PsycINFO) via OVID SP, using a search strategy comprising search terms specific to BD combined with a health economics search filter. Further searches using BD-specific search terms were performed in the Health Technology Assessment (HTA) database, the National Health Service [NHS] economic evaluation database (NHS EED) and the Central Register of Controlled Trials (CENTRAL) via Wiley. For both searches, a combination of key text words and medical subject headings (MeSH) was used. The databases were searched for articles published from January 1990 to 18 December 2015 (which was the date on which the searches were conducted). In addition, reference lists of identified systematic reviews of economic evaluations in the area of BD were scanned to identify further eligible studies.

The search strategy used in this study was adapted from the search strategy that was developed to inform the NICE (National Institute for Health and Care Excellence) updated guideline on BD in England and Wales [10]. Details on the search conducted for this study including the search strategy can be found in online Appendix 1.

## **1.2 Inclusion criteria**

The following inclusion criteria were applied:

- The study population comprised people of any age diagnosed with BD
- Interventions and comparators were any pharmacological treatment administered for the management of an acute episode (manic, depressive, mixed or rapid cycling) or as maintenance therapy for BD, whether licensed or

unlicensed for this particular indication, including pill placebo. The intervention needed to be described. Studies assessing treatments for BD without reference to specific interventions or studies assessing a mixture of different types of interventions, either as the intervention of interest or as the comparator, were excluded. Studies assessing the costs and consequences of different levels of treatment coverage were also excluded.

- Only full economic evaluations, assessing both costs and consequences, were considered; these included cost-utility analyses (CUAs), cost-effectiveness analyses (CEAs), cost-benefit analyses (CBAs) or cost consequence analyses (CCAs).
- Studies that considered exclusively harms of the interventions, without assessing their clinical benefits, that is, studies that did not assess the direct effect of interventions in treating BD, were not considered in the review. Studies which measured exclusively resource use and/or healthcare cost elements as a proxy to clinical outcomes were also not considered, as these were effectively cost analyses.
- The review considered studies based on decision-analytic modelling, as well as economic analyses conducted alongside single primary studies (clinical trials or cohort studies).
- Cost-of-illness studies, literature reviews and book chapters were excluded.
- Conference or dissertation abstracts, editorials, letters, commentaries and notes were excluded as they did not provide enough details for their methodological quality to be judged.

- Papers published from 1990 until the date of the search (18 December 2015) were considered. This date restriction was imposed so that retrieved economic evidence was more applicable to current healthcare settings and costs.
- Only papers in English language were included.

### **1.3 Study selection, data extraction, quality assessment and data synthesis**

All citations identified by the search were screened independently by the two authors. Citations that were clearly not relevant to the research question were excluded first, followed by independent screening of potentially relevant references against the inclusion criteria. Studies meeting inclusion criteria were agreed between the two authors; any disagreements were resolved by a third reviewer (ES). Data from studies that met inclusion criteria were extracted in a standardised data extraction form by one author (IM) and checked independently by the other author (JL). The following information was sought from each article: study population, interventions and comparators, study design, setting/country where the study was conducted, perspective, time horizon, primary measure(s) of outcome, sources of clinical, cost and (where applicable) utility data, cost elements included, currency and cost year, discounting rates, main events considered in modelling studies, methods for handling uncertainty, funding source, and cost effectiveness results (costs and outcomes where these were not synthesised, cases of dominance or Incremental Cost Effectiveness Ratios [ICERs]), including findings of statistical uncertainty and results of sensitivity analysis (such as estimates from Cost Effectiveness Acceptability Curves [CEACs]).

The methodological quality of the studies included in the review was assessed independently by the two authors using the Quality of Health Economic Studies (QHES) checklist [11]; any disagreements were resolved by a third reviewer (ES). The QHES is a 16-item checklist assessing the study design, the appropriateness and quality of clinical and cost evidence utilised, the method of analysis and the interpretation of the results. Each item on the checklist can be scored with a 'yes' or 'no' and is assigned a weighted point value, depending on its perceived relative importance. A study either meets or fails to meet each criterion, thus scoring either the full weighted value or zero for each question. The quality score for a study is calculated by adding up all of the points for questions answered 'yes'. Study quality scores can range between the best possible score of 100 (if all items are scored with a 'yes') and the lowest possible score of 0. The QHES checklist is provided in online Appendix 2.

The QHES checklist has been shown to be valid and reliable [12] and was selected among other checklists because its weighted scoring system allows comparisons across studies included in the review while taking into account the relative importance of appraisal criteria. Based on the obtained QHES checklist score, studies were classified as being of extremely poor quality (score 0-24); poor quality (score 25-49); fair quality (score 50-74); and high quality (score 75-100), a quality classification proposed by the checklist's developers [13] and adopted by other systematic reviews of economic evaluations [14-16]. If a QHES item was not applicable to a particular study, then the item was excluded from scoring and the study's score was multiplied by 100 and divided by the total sum of points of the

remaining items, producing an adapted score that was comparable to the other studies included in the review.

Results from individual studies were combined in a narrative synthesis; standard meta-analytic methods were deemed inappropriate as economic studies, in particular model-based, are themselves syntheses that may differ substantially in terms of study design, settings, costs and outcomes considered [17].

## **2. Results**

### **2.1 Results of the systematic search**

The systematic search of the literature retrieved 4,816 references (after duplicates were removed) published between January 1990 and 18 December 2015, of which 111 were potentially eligible for the review. Of these, 84 were excluded after reading the title and/or abstract and another 9 were excluded after reading the full paper as they did not meet inclusion criteria. Finally, 18 publications that met inclusion criteria were included in this review. The flow diagram of the search process is shown in Figure 1. A list of excluded studies with reasons for exclusion is provided in online Appendix 3.

### **2.2 Characteristics of included studies – overview of methodology**

The 18 publications included in the review reported data on 19 studies (one publication evaluated interventions at 2 different phases of BD) [18]. Of these, 7 studies assessed the cost effectiveness of pharmacological interventions for the treatment of manic, hypomanic or mixed episodes and/or rapid cycling in adults [19-24] or young people [25] with BD; 2 studies evaluated drugs for the management of

bipolar depression in adults [18;26]; 8 studies explored the cost effectiveness of drugs in the long-term (maintenance) treatment of BD in adults [18;27-33]; and 2 studies were economic evaluations of interventions for the management of patients in any phase of BD [34;35].

The majority of studies were conducted in the UK (8 studies in 7 publications) [18;19;24;25;28;29;31] or the US (7 studies) [22;23;26;27;30;32;33], with one study conducted in Italy [20], one study in the Netherlands [21] and 2 studies assessing the cost effectiveness of drugs for BD across different world sub-regions, as defined by WHO [34;35]. In their base-case analysis, all studies adopted a healthcare provider ( $\pm$  personal social services) [18-21;24;25;28;29;31;34;35] or a 3<sup>rd</sup> party payer [22;23;26;27;30;32;33] perspective and included direct medical costs. Only two studies considered a societal perspective (including indirect productivity losses) in a secondary or sensitivity analysis [18;33].

Three of the studies were CCAs that used various clinical and HRQoL outcome measures without attempting to synthesise costs and benefits [22;23;30]. Four studies were CEAs that used only natural outcome measures such as the percentage of people responding to treatment or remitting [19;21;27], the number of acute episodes avoided [29], the number of side effects avoided [21], or the mean length of hospital stay [21]. Twelve studies (11 publications) were CUAs that used the QALY [18;20;24;25;27;28;31-33] or the DALY [34;35] as the outcome measure, either exclusively or in addition to other natural outcome measures such as the number of acute episodes avoided [27;28;32;33], the number of euthymic days

achieved [27], or the percentage of people hospitalised due to acute episodes [28;32;33].

With the exception of 3 economic studies that were undertaken alongside RCTs [22;23;30], all other economic evaluations employed decision-analytic economic modelling techniques: in the majority of model-based studies (9) the authors constructed cohort Markov models [25;27-29;31-33] or hybrid models comprising decision-trees followed by cohort Markov model components [20;24], reflecting the cyclic nature of BD; 2 studies reported use of epidemiological mathematical models that appear to comprise individual-based Markov models [34;35]. Three studies (in 2 publications) adopted Discrete Event Simulation (DES) techniques [18;21]. Only 2 studies were based on simple decision-trees [19;26].

The health states and events considered across models were similar and reflected the course of BD: decision-trees (or decision-tree components) [19;20;24;26] and one DES model [21] considered the progression of acute episodes as reflected in one or more of the events of response, remission, hospital discharge, relapse to an acute episode and/or death. The main states considered in Markov models [20;24;25;27-29;31-35] and the other DES model [18] were euthymia and recurrence of acute episodes (manic, depressive, and, to a lesser degree, mixed), with some models incorporating additional events such as compliance with treatment, treatment discontinuation and drug sequencing, treatment resistance, and/or development of side effects (EPS and weight gain). Of the 16 models included in the review, 10 considered the risk of mortality and included the state of death [18;20;21;24;25;31-35].

All studies used RCTs as the source of the main efficacy data (i.e. response to acute treatment or relapse following maintenance treatment), in some cases supplemented with other longitudinal data and further assumptions. One study that was conducted alongside a 3-week RCT compared longer-term open label extension costs and outcomes with pre-randomisation data, effectively adopting a before-after design [22]. Only 2 model-based studies explicitly reported identification of RCTs via a systematic research of the literature [19;31].

The methods used for evidence synthesis differed across modelling studies. One study assumed that efficacy was the same across all interventions assessed, following inspection of RCT data [21]. Where available, studies that compared 2 interventions used (pooled) data directly from head-to-head comparisons between the interventions considered [20;24;28;29;32]. Where drugs had not been previously assessed in head-to-head RCTs, or where studies compared more than 2 drug treatments, RCT evidence was synthesised in a number of ways: two studies reported a narrative synthesis of rates reported in RCTs and other naturalistic studies, further adjusted by the authors [34;35]; four studies made indirect comparisons using placebo or lithium as baseline comparator [18;26;27;33]. Only 3 studies comparing multiple interventions employed network meta-analysis (NMA) for evidence synthesis [19;25;31].

All studies considered the main medical cost elements of BD treatment such as drug acquisition, laboratory testing, inpatient, outpatient, primary and community care (as appropriate, according to the settings treatment took place). A few modelling studies

[18;20;21;24;25] and, most likely, the RCT-based studies [22;23;30] included, implicitly or explicitly, the costs of management of side effects. All studies used resource use estimates that were mostly based on expert opinion, derived from either published expert estimates, physician surveys or clinical panels and further assumptions. Few studies utilised resource use data other than expert estimates, which were derived from RCTs [23;30], an RCT open label extension [22] and a chart review [29] and were all supplemented, to a varying degree, with expert opinion estimates. All studies used national unit costs supplemented, in a few studies, with local costs and other published prices.

The time horizon varied widely across the studies, ranging from 3 weeks [19] to lifetime [31;34;35]. In studies that examined treatments for acute manic or mixed episodes, or episodes of rapid cycling, the time horizon ranged from 3 weeks [19] to 5 years and 9 weeks [20;24]. In studies assessing drugs for the management of acute depression the time horizon varied between 3 months [26] and 5 years [18]. Studies assessing maintenance treatments and treatments over all phases of BD had longer time horizons, ranging from 1 year [29;30] to lifetime [31;34;35]. All studies with duration longer than one year used discounting for their costs and outcomes, except one study with a time horizon of 18 months in which the impact of discounting would most likely be negligible [27]. All other studies used a similar discount rate, either 3% [32-35] or 3.5% [18;20;24;25;28;31] annually.

All modelling studies undertook sensitivity analysis to explore the impact of the uncertainty characterising the input values and of alternative modelling scenarios on the base-case results. The majority of modelling studies employed Probabilistic

Sensitivity Analysis (PSA) to explore uncertainty [18-20;24-26;28;31-33;35]. The three CCAs conducted alongside RCTs [22;23;30] examined the level of statistical significance of the differences in costs and outcomes between interventions. However, the study samples were small so it is unlikely that the studies were powered to identify true differences in costs (and, potentially, in outcomes).

The vast majority of studies (14 studies in 13 publications, i.e. 74% of the studies) were funded by industry [18;20-22;22-24;26-30;32;33]; one further study was an industry submission to NICE in England, which was subsequently critically reviewed and supplemented with extra analyses by an independent Evidence Review Group on behalf of NICE [25]. Two studies were funded by WHO [34;35] and 2 studies were funded by the Health Technology Assessment (HTA) programme in the UK [19;31]. All studies enclosed a statement of funding.

### **2.3 Findings reported in the economic studies included in the systematic review**

The studies included in this systematic review suggest that, overall, pharmacological interventions are cost-effective compared with no treatment. Results and conclusions on the relative cost effectiveness between drugs were variable, depending also on the range of interventions assessed in each study and the cost effectiveness thresholds considered in each setting.

In the treatment of acute mania in adults, the 2 CCA studies conducted alongside RCTs suggested that olanzapine was cost-effective compared with no treatment (placebo) [22] and had similar costs and benefits with valproate [23] in the US. A UK

model-based study suggested that olanzapine was the most cost-effective drug if willingness to pay (WTP) exceeded £7,179/additional responder (2002 prices); otherwise haloperidol was the most cost-effective option. Other options such as valproate, lithium and quetiapine were all dominated by haloperidol and/or olanzapine [19]. The combination of risperidone/lithium was cost-effective among a range of options in the Netherlands. The combination of quetiapine/lithium was more effective than risperidone/lithium at an additional cost of €12,667/ responder (2003 prices) [21]. In the treatment of mixed episodes in adults, asenapine was more cost-effective than olanzapine in Italy [20] and the UK [24]. Aripiprazole was cost-effective as part of a strategy for the management of manic or mixed episodes in young people in the UK [25].

In the treatment of acute bipolar depression in adults, quetiapine alone or combined with a mood stabiliser (lithium or valproate) was found to be the most cost-effective option among a wide range of treatment strategies in the UK [18]. On the other hand, lurasidone was more effective than quetiapine XR in the US, at an additional cost of \$3,474/extra remission achieved (likely 2013 prices) [26].

In the maintenance treatment of adults with BD, lithium was found to have similar costs and benefits with valproate in the US, in a CCA conducted alongside a RCT [30]. One modelling study suggested that, at the lower cost effectiveness threshold of £20,000/QALY set by NICE in England, lithium was the most cost-effective strategy in the UK for people whose previous acute episode was manic (ICER versus olanzapine £11,359/QALY in 2005 prices), and also marginally the most cost-effective option in people whose previous acute episode was depressive (ICER of

lithium/imipramine versus lithium £21,370/QALY) [31]. However, in another modelling UK study in people who had recovered from a manic episode, lithium was found to be dominated by olanzapine [29]. In the US, lithium was found to have a probability of being cost-effective of 50% among a range of options, at a WTP of \$100,000/QALY [33]. In another modelling study conducted in the US olanzapine was dominated by lamotrigine; the latter had an ICER versus lithium reaching \$26,000/QALY (2004 prices) [27]. The combination of quetiapine or quetiapine XR with a mood stabiliser (lithium or valproate) was found to be cost-effective among a range of treatment strategies in 2 UK [18;28] and 2 US studies [32;33]. Two studies comparing lithium with valproate across all phases of the disease found lithium to be the dominant option [34;35].

Overall, results were found to be sensitive to the relative efficacy between drugs [24-27;33], the risk, length and cost of hospitalisation of acute episodes, in particular manic ones [27-29;32;33], the drug doses and drug acquisition costs [18;26-28;32;33], the consideration of productivity losses [18;33], the time horizon and/or discount rate [18;29], and the preventative effect of lithium on suicide [31].

#### **2.4 Overall study methodological quality**

According to their scoring on the QHES instrument, 6 studies were classified as of high quality (score 75-100) [19;20;24-26;32], 11 studies (in 10 publications) had fair quality (score 50-74) [18;22;27-31;33-35] and 2 studies poor quality (score 25-49) [21;23]. The vast majority of studies reported their objective, the study methods and study limitations in a clear way. All modelling studies described the model structure and main assumptions, although the rationale for the choice of the model type was

not always reported. All studies included a funding statement. The majority of studies used valid outcome measures. However, less than half of the studies used an appropriate method for clinical evidence synthesis, and therefore the quality of efficacy parameters was judged to be poor (even though the primary data source was RCTs). All modelling studies based most/all their resource use estimates, directly or indirectly (via published estimates), on expert opinion.

An overview of the methods employed by each study included in the systematic review is presented in Table 1. Results of each study, listed by measure of outcome used, both in their original currency and after their conversion to 2016 US\$ are provided in Table 2. A summary of each study's score on the QHES is shown in Table 3, with further notes on each study's strengths and limitations being provided in online Appendix 4.

### **3. Discussion**

#### **3.1 Overview of the systematic review findings**

The findings of this review indicate that economic evaluations in the area of BD have assessed the cost effectiveness of a wide range of pharmacological treatments including lithium, antiepileptic drugs used as mood stabilisers (carbamazepine, lamotrigine, valproate), older and newer antipsychotics (haloperidol, aripiprazole, asenapine, lurasidone, olanzapine, quetiapine, risperidone), antidepressants (imipramine) and combinations of some of the drugs. The results suggest that pharmacological interventions are cost-effective compared with no treatment in the management of BD and that antipsychotic drugs alone or combined with lithium or valproate are likely to be the most cost-effective options in the pharmacological

treatment of acute manic, mixed or depressive episodes. Lithium appears to be more cost-effective as a maintenance treatment, whereas lamotrigine, olanzapine and quetiapine combined with a mood stabiliser are alternative cost-effective options in the maintenance treatment of BD. However, these conclusions are determined by the availability of data on each phase of the disorder, and the focus of research on specific drugs and comparisons, which appears to have been driven by marketing purposes, as 74% of the studies were funded by industry. It needs to be noted that, in addition to the limitation of focusing on specific drugs and making limited comparisons, studies sponsored by drug manufacturers entail potential conflicts of interest and may be biased in the conduct and interpretation of findings.

## **3.2 Critique on methods adopted by the economic studies**

### **3.2.1 Interventions and comparisons**

The relative cost effectiveness of an intervention is determined by the number and range of alternative options considered in the comparison. More than 50% (11/19) of the studies included in the review [20;22-24;26;28-30;32;34;35] assessed the relative cost effectiveness between 2 pharmacological interventions only, ignoring a wide range of other drugs that are effective in the treatment of BD and would thus comprise appropriate comparators. Therefore, the results of these studies are of limited value. Some studies comparing more than 2 interventions made selective pairwise comparisons between one intervention of interest and its comparators, instead of conducting appropriate incremental analysis of all treatment options considered, either in the base-case or in sensitivity analysis [18;21]. This can lead to results reported being potentially misleading as they exaggerate the magnitude of the cost effectiveness of the intervention of interest. Only few studies compared a

wide range of interventions and undertook appropriate incremental analysis in both base-case and sensitivity analysis [19;25;27;31;33].

### **3.2.2 Types of economic evaluation and outcome measures used**

Interpretation of the results reported in each study was straightforward in cases of dominance, regardless of the outcome measure used, and in CUAs, in settings where a cost effectiveness threshold has been determined, such as the £20,000-£30,000/QALY threshold set by NICE in England [36]. In all other cases where natural outcome measures were used, results are not easy to interpret and assessment of cost effectiveness requires judgements on whether the incremental benefit is worth the extra cost incurred for its achievement, which is related to the decision-makers' WTP. Furthermore, use of different outcome measures across studies makes comparisons of cost effectiveness practically impossible. The 3 studies conducted alongside RCTs were CCAs that did not aggregate costs and outcomes, and therefore interpretation of cost effectiveness was also problematic [22;23;30].

Within CUAs, comparability of QALY outcomes is limited if different utility measures have been adopted; for example, it has been shown that there are discrepancies between utility changes measured using EQ-5D and SF-6D in mental health populations [37] and other patient groups [38]. The 10 CUAs included in this review that used the QALY as the measure of outcome obtained utility data from a range of studies that adopted different methodologies for elicitation of utility values. Six of the CUAs (in 5 publications) [18;20;24;25;31] used mainly published utility data derived from two studies: Revicki et al. [39], who used vignettes to describe hypothetical,

drug-specific health states for mania, severe bipolar depression and stable BD that were valued by stable outpatients with BD in the US; and Hayhurst et al. [40], who estimated utility values for the BD-related states of euthymia and depression using EQ-5D ratings from participants in a multicentre RCT of cognitive behavioural therapy and the EQ-5D-3L UK tariff [41]. Additional utility data used in these CUAs included dis-utilities for side effects such as EPS and weight gain that were taken from studies in schizophrenia [42-44] and adolescents [45], EQ-5D-based utilities specific to unipolar depression [46], and other unpublished or further adjusted utility data. The remaining 4 CUAs [27;28;32;33] estimated utility values for health states of BD using unpublished SF-36 data obtained from participants in lamotrigine pivotal trials and the SF-6D utility algorithm for the UK population [47].

Several regulatory and advisory bodies worldwide advocate the use of generic preference-based measures (such as the EQ-5D and the SF-6D) for the estimation of QALYs in CUAs of healthcare interventions, to enhance comparability across CUAs; for example, NICE in England has explicitly expressed a preference for the use of EQ-5D

[36]. On the other hand, the evidence on the appropriateness of the use of generic HRQoL measures (such as EQ-5D and SF-36) in BD is mixed, with concerns being expressed around how to obtain reliable HRQoL ratings from patients in manic or hypomanic states [48]. It has to be noted that utility data for manic states were derived, in the majority of economic studies included in the review, from stable patients with BD who expressed their preferences for vignettes [39]. Nevertheless, the literature around the comparability of different utility measures and the measurement of preferences in patients with BD suggest that the QALY results

obtained from the CUAs included in the review should be interpreted with caution regarding their comparability and, possibly, appropriateness.

### **3.2.3 Settings, perspectives and cost elements**

Although all studies adopted a healthcare provider or 3<sup>rd</sup> party payer perspective and included direct medical costs exclusively or in their base-case analysis, it is unlikely that they are comparable in this aspect: the studies were carried out in a variety of countries including the UK [18;19;24;25;28;29;31], the US [22;23;26;27;30;32;33], Italy [20], the Netherlands [21] as well as a number of world sub-regions, as defined by WHO [34;35]. The funding and structure of healthcare services, access and BD treatment pathways (and therefore related resource use) and unit costs are likely to differ across countries, thus limiting the comparability of the results and their transferability from one setting to another. The comparability of study cost results is further limited by the use of different price years and currency across studies, although appropriate adjustments are possible to bring all cost figures to the same year and currency of interest. Table 2 presents the results of all studies in 2016 US dollars, in addition to the original currency and cost year used in the study; however, it needs to be emphasised that the comparability, and thus transferability, of these figures is limited if they are derived from different countries, for the reasons stated above.

An important consideration when interpreting the results of the review is whether drugs are patented or available in generic form. A number of the drugs used in BD that were assessed in the economic literature, such as quetiapine (but not quetiapine XR), olanzapine and lamotrigine are now available in generic form, and therefore

their acquisition cost is lower than the cost of the patented forms evaluated in the studies included in the systematic review. Thus their current cost effectiveness is likely higher than that reported in the economic studies considered in the review.

Another point to consider when interpreting cost effectiveness is the perspective adopted in the study: it has been argued that economic evaluations in health care should adopt a broad societal perspective [49]. Only two studies in the review considered a societal perspective (including indirect productivity losses) in a secondary or sensitivity analysis [18;33]. Adoption of a societal perspective is overall expected to improve the cost effectiveness of more effective drugs, as better health outcomes are likely to lead to improved productivity, lower unemployment rates, and less time off work. However, incorporation of productivity losses may have equity implications (as it disfavours elderly, children and people with disability) and raises methodological issues in terms of quantifying and valuing productivity effects [50;51].

### **3.2.4 Trial-based economic evaluations: study size, power calculations and missing data**

One important aspect of trial-based economic evaluations is whether the studies are sufficiently powered to detect differences in costs between interventions. Of the 3 trial-based studies, one compared costs and outcomes of olanzapine between a RCT open-label extension and the pre-randomisation period, effectively adopting the design of a controlled before-after study with all the associated limitations of this study design [52]. The other 2 trial-based analyses [23;30] were likely insufficiently powered to detect differences in costs; this is because the skewness in cost data requires greater sample sizes for cost comparisons compared with clinical outcome

comparisons, and usually power calculations in RCTs are based on clinical outcomes [53]. Therefore it is not surprising that both studies did not detect any statistically significant difference in costs between interventions assessed (they did not report any significant difference in outcomes either). Moreover, the studies reported moderate to high attrition rates (14-57%) without reporting methods for handling missing data. However, inappropriate methods for handling missing data can lead to misleading results [54].

### **3.2.5 Modelling studies: types of models, model structures, events and health states considered**

The majority of economic models were Markov models [25;27-29;31-33] or hybrid models comprising decision-trees followed by cohort Markov model components [20;24]; 2 studies were based on simple decision-trees [19;26]. Another 2 studies reported use of epidemiological mathematical models that appear to comprise individual-based Markov models [34;35]. Three studies (in 2 publications) adopted DES techniques [18;21].

Decision-trees may be appropriate for the simulation of a short-term period comprising a single pathway of events and health states in the course of the BD such as the development (and potential resolution) of an acute episode, although they are not suitable for the simulation of recurrent events over a long period of time [55;56]. Indeed, the studies based on decision-trees assessed the cost effectiveness of drugs in the management of a manic [19] or depressive [26] episode, and therefore, in principle, the type of the selected model appears to be appropriate.

The studies that used Markov models, alone or combined with decision-trees, modelled a sequence of repetitive events (such as remission and relapse to manic or depressive episodes) over a number of model cycles (with the cycle length being predefined and stable over time), by allowing people to transition between different, distinct health states; they were thus more appropriate for the assessment of longer-term costs and outcomes over the cyclic course of BD [57]. One characteristic of Markov models is the lack of 'memory' for earlier cycles; this means that the transition probabilities from one state to another depend only on the current health state and are independent from history of transitions / events that have occurred in previous cycles (Markovian assumption); this limitation, though, can be resolved with the incorporation in the model of temporary 'tunnel' states, which are arranged in a way so that each can only transition to the next state, keeping a memory of the history of events [56;57]. Nevertheless, all Markov models considered in this review used constant transition probabilities over time, suggesting that the annual risk of acute episodes (manic or depressive) in the populations considered in the models was independent of the number and nature of previous episodes. However, it has been found that the risk of relapse in people with BD increases with the number of previous episodes [58], suggesting that the existing Markov models in the area of BD are likely to have underestimated the cumulative risk of recurrence of acute episodes, and thus the total time patients spent in an acute episode over each model's time horizon. Ultimately, they may have underestimated the impact of maintenance treatments on the long-term course of BD and their relative cost effectiveness.

In contrast, DES models can simulate the behaviour of the study population at an individual level over time. Each individual is assigned personal characteristics defined at the start of the simulation and is allowed to move from one event to another with probabilities that are determined over time by these baseline characteristics, as well as by changes in these characteristics through events experienced by the individual within the model [56;59]. DES models can accommodate differing cycle lengths more easily than Markov models, are not limited by the Markovian assumption, and therefore may be more appropriate (though more time-consuming) for use in the area of BD [60].

The structure of all models appears to be appropriate for modelling the course of BD, although some studies adopted a rather simplistic approach and omitted some states and events associated with BD, in particular short-term studies that focused on the treatment and progression of acute episodes. One important limitation of a number of studies was the lack of consideration of side effects, and their impact on both costs (incurred for their management) and the patients' HRQoL. Such an omission may have potentially overestimated the cost effectiveness of drugs that are associated with serious, long-term side effects that reduce the patients' quality of life and are costly to treat. For example, atypical antipsychotics (olanzapine, risperidone, aripiprazole, etc.) can lead to the development of metabolic syndrome with an increased risk for diabetes, cardiovascular disease and their complications [61-64]; lithium is associated with increased risk of hypothyroidism, hyperparathyroidism, weight gain and reduced urinary concentrating ability, although the risk of end-stage renal failure is low [65].

People with BD are at increased mortality risk, in particular due to suicide, compared with the general population [66;67]. About 17% of people with BD type I and 24% of those with BD type II attempt suicide during the course of their illness [68]. Lithium is the only known drug in the treatment of bipolar disorder with a preventative effect against suicide [69]. Omission of the death state from BD economic models has been criticised [60], however, the 6 modelling studies that did not consider mortality had short time horizons ranging from 3 weeks to 2 years [19;26-29], and therefore it is unlikely that this omission has had a significant impact on their results. Moreover, given that the risk of mortality in people with BD appears to be equal across drug treatments with the exception of lithium, omission of the death state is likely to have affected results in an even lesser degree in the models that did not include lithium.

### **3.2.6 Efficacy data sources and methods for evidence synthesis**

The model-based studies included in the review utilised efficacy data from single RCTs [29] or (in most cases) synthesised data derived from a number of RCTs. However, only 2 model-based studies explicitly reported identification of RCTs via a systematic research of the literature [19;31]. Selective use of RCT efficacy data, instead of a systematic literature search for identification of clinical evidence and population of economic models, is likely to have introduced bias in the economic analyses. RCT data were synthesised either via a narrative synthesis [34;35], data pooling across RCTs of similar design [20;24;28;32;33], an indirect comparison (using a control as common comparator) [18;26;27;33] or a NMA [19;25;31].

NMA is the simultaneous comparison of multiple competing treatments in a single statistical model [70]. NMA enables evaluation of the relative effectiveness between

multiple interventions, even if they have not been compared in an individual trial, and allows simultaneous inference on all interventions by combining both direct and indirect evidence between interventions, while preserving randomisation [70-72]. This approach for evidence synthesis is essential for populating model-based economic studies assessing more than two competing interventions.

A key assumption when conducting NMA (or simple indirect comparisons using a common baseline comparator) is that the populations included in all RCTs are similar so that the treatment effects are exchangeable across all patients [71]. However, this was not the case in some of the studies that synthesised evidence via indirect comparisons or NMA: in Ekman et al. [18], the study populations in the RCTs included in the indirect comparisons were not similar in terms of the phase of the disorder. In 3 economic analyses of maintenance treatments [27;31;33] the synthesised evidence came from RCTs with different study designs and study populations: some RCTs assigned stable patients to a new maintenance treatment, while other trials randomised participants remitting from an acute episode (either manic or depressive) to continue or discontinue treatment that had been effective in the acute phase. However, patients that have already responded to one drug while in an acute episode are more likely to respond to this particular drug in the maintenance phase, rather than to other drugs included in the NMA, thus violating the assumption of exchangeability.

### **3.2.7 Time horizon and discounting**

BD is a chronic disorder with high risk of recurrence over time, and therefore long-term (maintenance) treatments for BD should be ideally modelled over lifetime, to

allow full consideration of the preventative effect of maintenance therapy against the cumulative risk of future recurrence of acute episodes over time. Shorter time horizons in studies assessing treatments for acute episodes could be appropriate, as long as they cover the full progression of the episode, from start of treatment through to resolution, allowing adequate follow-up time to assess the risk of relapse. Manic episodes usually last for between 2 weeks and 4-5 months (median duration about 4 months), whereas depressive episodes tend to last longer (median length about 6 months), though rarely for more than a year, except in the elderly [1]. Thus, studies assessing the cost effectiveness of drugs for the management of acute episodes that had shorter time horizons than the typical duration of the type of episode examined, as well as studies that did not allow enough time for the risk of relapse to be realised, have inadequately captured the costs and benefits associated with treatment.

### **3.2.8 Methods for handling uncertainty**

All model-based studies included, to a variable extent, sensitivity analyses to explore the robustness of the results under the uncertainty characterising input parameters and model structural assumptions. In deterministic sensitivity analysis, input parameters are assigned a range of point estimates to explore the impact of the change in estimates on the results of the analysis. In PSA, input parameters are assigned appropriate probabilistic distributions; subsequently, the model is run over a large number of iterations, each drawing random values out of the distributions fitted onto the model input parameters, resulting in the full uncertainty characterising all input parameters being concurrently considered in the base-case results; therefore, PSA is considered the most appropriate form of sensitivity analysis [73]. Results can then be presented in the form of CEACs, which show the probability of

each intervention being cost-effective at different levels of WTP per additional unit of benefit. This methodology can be used both in modelling studies and studies conducted alongside primary clinical studies [74;75]. Bootstrapping techniques can be used for conducting PSA in trial-based economic evaluations [76]. A large number of modelling studies included in the review employed PSA to explore the uncertainty in their results [18-20;24-26;28;31-33;35].

Conclusively, the studies included in the review are characterised by large differences in the interventions assessed and the comparisons made, the type of economic evaluations, the outcome measures used, the methods for evidence synthesis, the settings and the perspectives adopted, the modelling techniques and the time horizon. The quality of the studies was variable, however, 17 out of the 19 studies included in the review had a high or fair quality according to their QHES score.

### **3.3 Strengths and limitations of the systematic review**

The current review assessed, in a systematic way, all the available economic evidence on pharmacological interventions in the management of bipolar disorder. Searches were conducted in 6 electronic bibliographic databases, using a broad range of search terms. The systematic review was conducted by two researchers following golden standards for conducting systematic reviews [9] and utilised a validated checklist for the appraisal of economic studies [11]. The methods and results of each study were considered in the context of the phase of BD that was the focus of the pharmacological intervention (i.e. treatment of acute manic, mixed or depressive episodes or rapid cycling, or maintenance therapy in stable BD).

However, one limitation of the review is that it considered only English papers, and therefore papers potentially published in other languages have been omitted.

The QHES checklist that was used for the appraisal of the economic studies has been designed for model-based studies. However, it was possible to use for the 3 trial-based economic evaluations included in the review after omitting non-relevant items and adjusting the total checklist score. One limitation of the QHES checklist is that, for each dichotomous answer (yes/no), it gives the full or zero score, not allowing for intermediate scores, which would be more appropriate if a checklist item is partially satisfied. Moreover, judgements on whether criteria are met can be subjective.

A major limitation of using the QHES checklist (as well as other available methodological checklists) is that some items may be fundamental in determining the quality and credibility of the study, yet the only 'penalisation' of the study is getting a zero score on these items. For example, the source of efficacy data and the methods for evidence synthesis is one of the most important methodological 'characteristics' of a study; a biased method of evidence synthesis may have huge effects on the analysis, distorting extensively the results. Yet the QHES relevant checklist item on source of variable estimates has a score of only 8 (out of 100). Moreover, this item refers to a single source of efficacy data ("randomized control trial - best, expert opinion – worst"), and does not explicitly mention the method for evidence synthesis, which may be essential, in particular when a large number of competing interventions are assessed in a modelling study.

Although one of the aims of this review was to synthesise, in a narrative way, the available evidence in order to identify cost-effective pharmacological interventions for the treatment of BD, a robust conclusion on the relative cost effectiveness between interventions was not possible, due to differences across studies in terms of study design, interventions compared, and other factors described above, which limited the comparability of results across studies and did not allow for consistent conclusions to be drawn. Nevertheless, a clear message from the review findings is that pharmacological interventions are cost-effective compared with no treatment in the management of people with BD, and this applies to both treatment of the acute phase of the disorder, and also to the long-term prophylactic management.

### **3.4 Consideration points for researchers conducting economic research and for health professionals and policy makers interpreting published economic evaluations**

1. Economic evaluations in the area of BD should consider a wide range of treatment options, including standard care, in order to assess the relative cost effectiveness of interventions. Model-based economic evaluations should ideally consider all available treatment options, including psychological interventions, if relevant (this would not be feasible for economic analyses undertaken alongside primary clinical studies).
2. The perspective of the analysis should be clearly stated and justified and all relevant costs elements should be considered. Depending on the objective of the analysis, wider perspectives, beyond healthcare ones (e.g. public sector or societal perspective) may be appropriate to adopt, especially if there are relevant costs falling to other sectors (e.g. social care).

3. Economic studies should have long enough time horizons to capture the full impact of interventions on benefits, harms and costs; given the chronicity of BD, ideally time horizons of model-based studies should extend over lifetime. Trial-based economic evaluations should extrapolate findings beyond the time frame of the trial if possible, by employing decision-analytic modelling.
4. Generic utility measures such as EQ-5D and SF-36 may not be adequate in capturing HRQoL changes in people with BD. On the other hand, there are recommendations on the use of generic measures for the estimation of QALYs to enhance comparability across disease areas. Use of utility data elicited using BD-related vignettes or a BD-specific measure (currently not available) in sensitivity analysis, or use of a secondary (clinical) outcome measure might be alternative options.
5. Evidence on the efficacy in model-based studies should be derived from a systematic review of the literature and synthesised in an appropriate, methodologically robust, way. NMA is the optimal methodology for evidence synthesis when more than two interventions are compared, as it enables consideration of all available evidence coming from direct and indirect comparisons, without ignoring any part of the evidence and while preserving randomisation.
6. The type of economic model should be appropriate for the phase of the disorder examined. Decision-trees may be adequate for the evaluation of drugs focusing on treatment of acute BD episodes. Markov models that employ tunnel states to keep a memory of the history of events and DES models are more appropriate for the assessment of the long-term cost effectiveness of interventions for BD, in particular as maintenance treatments.

7. Any economic model should consider relevant health states, events and outcomes associated with BD and the evaluated interventions. Euthymic and acute manic or depressive states are essential states in any BD model, reflecting the cyclic nature of BD. Chronic side effects that may have a prolonged impact on costs and HRQoL should not be omitted. Mortality may also be an important parameter, especially in models with long time horizons and in evaluations of lithium, due to the latter's effect in preventing suicide.
8. Extensive deterministic and probabilistic sensitivity analysis should be employed to explore the robustness of the results under the uncertainty characterising input parameters and under alternative scenarios.
9. Regarding the generalisability of results and their transferability across different countries/settings, consideration should be given to the similarities and differences between the settings in terms of the study population characteristics, the funding of the services for people with BD, the treatment pathways, the resource use and implementation issues associated with each intervention, and the unit costs. The perspective of the analysis should also be assessed against the context of interest. A point to note is that as drugs become generic over time, their cost effectiveness eventually improves.
10. Conclusions on cost effectiveness may be difficult to reach if an intervention is not clearly dominant over another; judgements on cost effectiveness ultimately rely on the cost effectiveness threshold adopted, which depends on the policy makers' WTP for treatment benefits, which may vary across countries and health systems.

#### **4. Conclusion**

Available evidence suggests that pharmacological interventions are cost-effective compared with no treatment in the management of people with BD and should be used in the management of both the acute and the maintenance phase of the disorder; however, it is difficult to draw any robust conclusions on the relative cost effectiveness between drugs, due to differences in the range of interventions assessed, the methods employed, and the settings across economic studies, as well as methodological caveats characterising many of them. Future economic evaluations need to consider the whole range of treatment options available for the management of BD and adopt appropriate methodology, especially for evidence synthesis, addressing the limitations of existing literature, to explore more robustly the relative cost effectiveness of pharmacological interventions in the area of BD.

**Compliance with ethical standards**

**Funding:** no funding has been received for this study

**Conflict of interest:** The authors declare to have no conflict of interest.

**Ethical approval:** This article does not contain any studies with human participants performed by the authors.

**Informed consent:** Not relevant

**Acknowledgement:** the authors wish to thank Eric Slade, who agreed to act as a third reviewer to resolve any disagreements between the authors regarding the selection of eligible studies for the review and the quality assessment of included studies.

## References

1. World Health Organization. The ICD-10 Classification of Mental and Behavioural Disorders. Clinical Descriptions and Diagnostic Guidelines. 1992. World Health Organization.

Ref Type: Report

2. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders (Fifth Ed.). 2013. Arlington, Va, American Psychiatric Publishing.

Ref Type: Report

3. Murray CJ, Vos T, Lozano R, Naghavi M, Flaxman AD, Michaud C et al. Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet* 2012; 380(9859):2197-2223.

4. 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.

5. Runge C, Grunze H. Annual costs of bipolar disorders in Germany. *Nervenarzt* 2004; 75(9):896-903.

6. Hakkaart-van Roijen L, Hoeijenbos MB, Regeer EJ, ten Have M, Nolen WA, Veraart CP et al. The societal costs and quality of life of patients suffering from bipolar disorder in the Netherlands. *Acta Psychiatr Scand* 2004; 110(5):383-392.

7. Dilsaver SC. An estimate of the minimum economic burden of bipolar I and II disorders in the United States: 2009. *Journal of Affective Disorders* 2011; 129:79-83.

8. National Institute for Health and Care Excellence. Bipolar Disorder: assessment and management. NICE clinical guideline 185. 2014. London, NICE.

Ref Type: Report

9. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gotzsche PC, Ioannidis JP et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *BMJ* 2009; 339:b2700.

10. National Collaborating Centre for Mental Health. Bipolar disorder: The assessment and management of bipolar disorder in adults, children and young people, in primary and secondary care. Commissioned by the National Institute for Health and Care Excellence. 2014. Leicester & London: The British Psychological Society and the Royal College of Psychiatrists.

Ref Type: Report

11. Ofman JJ, Sullivan SD, Neumann PJ, Chiou CF, Henning JM, Wade SW et al. Examining the value and quality of health economic analyses: implications of utilizing the QHES. *J Manag Care Pharm* 2003; 9(1):53-61.
12. Walker DG, Wilson RF, Sharma R, Bridges J, Niessen L, Bass EB et al. Best practices for Conducting Economic Evaluations in Health Care: A systematic Review of Quality Assessment Tools. Methods Research Report (prepared by John Hopkins University Evidence-based Practice Center under contract No. 290-2007-10061-I) AHRQ Publication No. 12(13)-EHC132-EF. 2012. Rockville, MD, Agency for Healthcare Research and Quality.

Ref Type: Report

13. Chiou CF, Hay JW, Wallace JF, Bloom BS, Neumann PJ, Sullivan SD et al. Development and validation of a grading system for the quality of cost-effectiveness studies. *Med Care* 2003; 41(1):32-44.
14. Spiegel BM, Targownik LE, Kanwal F, Derosa V, Dulai GS, Gralnek IM et al. The quality of published health economic analyses in digestive diseases: a systematic review and quantitative appraisal. *Gastroenterology* 2004; 127(2):403-411.
15. Foster WJ, Tufail W, Issa AM. The quality of pharmacoeconomic evaluations of age-related macular degeneration therapeutics: a systematic review and quantitative appraisal of the evidence. *Br J Ophthalmol* 2010; 94(9):1118-1126.
16. Jones M, Lewis S, Parrott S, Coleman T. Systematic critical review of previous economic evaluations of smoking cessation during pregnancy. *BMJ Open* 2015; 5(11):e008998.
17. Pignone M, Saha S, Hoerger T, Lohr KN, Teutsch S, Mandelblatt J. Challenges in systematic reviews of economic analyses. *Ann Intern Med* 2005; 142(12 Pt 2):1073-1079.
18. Ekman M, Lindgren P, Miltenburger C, Meier G, Locklear J. Cost effectiveness of quetiapine in patients with acute bipolar depression and in maintenance treatment after an acute depressive episode. [References]. *Pharmacoeconomics* 2012; 30(6):513-530.
19. Bridle C, Palmer S, Bagnall AM, Darba J, Duffy S, Sculpher M et al. A rapid and systematic review and economic evaluation of the clinical and cost-effectiveness of newer drugs for treatment of mania associated with bipolar affective disorder. *Health Technology Assessment* 2004; 8(19):iii-109.
20. Caresano C, Di SG, Fagiolini A, Maina G, Perugi G, Ripellino C et al. Cost-effectiveness of asenapine in the treatment of patients with bipolar I disorder with mixed episodes in an Italian context. *Adv Ther* 2014; 31(8):873-890.
21. Klok RM, Al Hadithy AF, van Schayk NP, Antonisse AJ, Caro JJ, Brouwers JR et al. Pharmacoeconomics of quetiapine for the management of acute mania

in bipolar I disorder. *Expert Review of Pharmacoeconomics and Outcomes Research* 2007; 7(5):459-467.

22. Namjoshi MA, Namjoshi MA. Economic, clinical, and quality-of-life outcomes associated with olanzapine treatment in mania. Results from a randomized controlled trial. *Journal of Affective Disorders* 2002; 69(1-3):109-118.
23. Revicki DA, Paramore LC, Sommerville KW, Swann AC, Zajecja JM, Depakote Comparator Study Group. Divalproex sodium versus olanzapine in the treatment of acute mania in bipolar disorder: health-related quality of life and medical cost outcomes. *Journal of Clinical Psychiatry* 2003; 64(3):288-294.
24. Sawyer L, Azorin JM, Chang S, Rinciog C, Guiraud-Diawara A, Marre C et al. Cost-effectiveness of asenapine in the treatment of bipolar I disorder patients with mixed episodes. *Journal of Medical Economics* 2014; 17(7):508-519.
25. Uttley L, Kearns B, Ren S, Stevenson M. Aripiprazole for the treatment and prevention of acute manic and mixed episodes in bipolar I disorder in children and adolescents: a NICE single technology appraisal. *Pharmacoeconomics* 2013; 31(11):981-990.
26. Rajagopalan K, Meyer K, O'Day K, Denno M, Loebel A. Cost-effectiveness of lurasidone vs quetiapine extended-release (XR) in patients with bipolar depression. *Journal of Medical Economics* 2015; 18(10):821-827.
27. Calvert NW, Burch SP, Fu AZ, Reeves P, Thomson TR. The cost-effectiveness of lamotrigine in the maintenance treatment of adults with bipolar I disorder. *Journal of Managed Care Pharmacy* 2006; 12(4):322-330.
28. Fajutrao L, Paulsson B, Liu S, Locklear J. Cost-effectiveness of quetiapine plus mood stabilizers compared with mood stabilizers alone in the maintenance therapy of bipolar I disorder: results of a Markov model analysis. *Clinical Therapeutics* 2009; 31 Pt 1:1456-1468.
29. McKendrick J, Cerri KH, Lloyd A, D'Ausilio A, Dando S, Chinn C. Cost effectiveness of olanzapine in prevention of affective episodes in bipolar disorder in the United Kingdom. *Journal of Psychopharmacology* 2007; 21(6):588-596.
30. Revicki DA, Hirschfeld RM, Aheam EP, Weisler RH, Palmer C, Keck PE Jr. Effectiveness and medical costs of divalproex versus lithium in the treatment of bipolar disorder: results of a naturalistic clinical trial. *Journal of Affective Disorders* 2005; 86(2-3):183-193.
31. Soares-Weiser K, Bravo Vergel Y, Beynon S, Dunn G, Barbieri M, Duffy S et al. A systematic review and economic model of the clinical effectiveness and cost-effectiveness of interventions for preventing relapse in people with bipolar disorder. *Health technology assessment (Winchester, England)* 2007; 11(39):iii-iv.

32. Woodward TC, Tafesse E, Quon P, Kim J, Lazarus A. Cost-effectiveness of quetiapine with lithium or divalproex for maintenance treatment of bipolar I disorder. *Journal of Medical Economics* 2009; 12(4):259-268.
33. Woodward TC, Tafesse E, Quon P, Lazarus A. Cost effectiveness of adjunctive quetiapine fumarate extended-release tablets with mood stabilizers in the maintenance treatment of bipolar I disorder. *PharmacoEconomics* 2010; 28(9):751-764.
34. Chisholm D, van Ommeren M, Ayuso-Mateos JL, Saxena S. Cost-effectiveness of clinical interventions for reducing the global burden of bipolar disorder. *British Journal of Psychiatry* 2005; 187:559-567.
35. Chisholm D, Saxena S. Cost effectiveness of strategies to combat neuropsychiatric conditions in sub-Saharan Africa and South East Asia: mathematical modelling study. *BMJ* 2012; 344:e609.
36. National Institute for Health and Care Excellence. Process and methods guides. Guide to the Methods of Technology Appraisal. London: National Institute for Health and Care Excellence; 2013.
37. Lamers LM, Bouwmans CA, van SA, Donker MC, Hakkaart L. Comparison of EQ-5D and SF-6D utilities in mental health patients. *Health Econ* 2006; 15(11):1229-1236.
38. Brazier J, Roberts J, Tsuchiya A, Busschbach J. A comparison of the EQ-5D and SF-6D across seven patient groups. *Health Econ* 2004; 13(9):873-884.
39. Revicki DA, Hanlon J, Martin S, Gyulai L, Nassir Ghaemi S, Lynch F et al. Patient-based utilities for bipolar disorder-related health states. *J Affect Disord* 2005; 87(2-3):203-210.
40. Hayhurst H, Palmer S, Abbott R, Johnson T, Scott J. Measuring health-related quality of life in bipolar disorder: relationship of the EuroQol (EQ-5D) to condition-specific measures. *Qual Life Res* 2006; 15(7):1271-1280.
41. Dolan P. Modeling valuations for EuroQol health states. *Med Care* 1997; 35(11):1095-1108.
42. Heeg B, Buskens E, Botteman M, Caleo S, Ingham M, Damen J et al. The cost-effectiveness of atypicals in the UK. *Value Health* 2008; 11(7):1007-1021.
43. Briggs A, Wild D, Lees M, Reaney M, Dursun S, Parry D et al. Impact of schizophrenia and schizophrenia treatment-related adverse events on quality of life: direct utility elicitation. *Health Qual Life Outcomes* 2008; 6:105.
44. Mangalore R. The utility approach to valuing health states in schizophrenia. *Mental Health Research Review* 2000; 7:11-15.
45. Boyle SE, Jones GL, Walters SJ. Physical activity, quality of life, weight status and diet in adolescents. *Qual Life Res* 2010; 19(7):943-954.

46. Sobocki P, Ekman M, Agren H, Krakau I, Runeson B, Martensson B et al. Health-related quality of life measured with EQ-5D in patients treated for depression in primary care. *Value Health* 2007; 10(2):153-160.
47. Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. *J Health Econ* 2002; 21(2):271-292.
48. Brazier J, Connell J, Papaioannou D, Mukuria C, Mulhern B, Peasgood T et al. A systematic review, psychometric analysis and qualitative assessment of generic preference-based measures of health in mental health populations and the estimation of mapping functions from widely used specific measures. *Health Technol Assess* 2014; 18(34):vii-xxv, 1.
49. Jonsson B. Ten arguments for a societal perspective in the economic evaluation of medical innovations. *Eur J Health Econ* 2009; 10(4):357-359.
50. Sculpher M. The role and estimation of productivity costs in economic evaluation. In: Drummond M, McGuire A, editors. *Economic evaluation in health care. Merging theory with practice*. New York: OUP; 2001. p. 94-112.
51. Knies S, Severens JL, Ament AJ, Evers SM. The transferability of valuing lost productivity across jurisdictions. differences between national pharmacoeconomic guidelines. *Value Health* 2010; 13(5):519-527.
52. *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]*. Available from [www.cochrane-handbook.org](http://www.cochrane-handbook.org). Higgins JP, Green S, editors. 2011. The Cochrane Collaboration.

Ref Type: Report

53. Briggs A. Economic evaluation and clinical trials: size matters. *BMJ* 2000; 321(7273):1362-1363.
54. Faria R, Gomes M, Epstein D, White IR. A guide to handling missing data in cost-effectiveness analysis conducted within randomised controlled trials. *Pharmacoeconomics* 2014; 32(12):1157-1170.
55. Barton P, Bryan S, Robinson S. Modelling in the economic evaluation of health care: selecting the appropriate approach. *Journal of Health Services Research and Policy* 2004; 9:110-118.
56. Petrou S, Gray A. Economic evaluation using decision analytical modelling: design, conduct, analysis, and reporting. *BMJ* 2011; 342:d1766.
57. Sonnenberg FA, Beck JB. Markov Models in Medical Decision Making: A Practical Guide. *Medical Decision Making* 1993; 13:322-338.
58. Kessing LV, Hansen MG, Andersen PK, Angst J. The predictive effect of episodes on the risk of recurrence in depressive and bipolar disorders - a life-long perspective. *Acta Psychiatr Scand* 2004; 109(5):339-344.
59. Caro JJ. Pharmacoeconomic analyses using discrete event simulation. *Pharmacoeconomics* 2005; 23:323-332.

60. Mohiuddin S. A systematic and critical review of model-based economic evaluations of pharmacotherapeutics in patients with bipolar disorder. *Applied Health Economics and Health Policy* 2014; 12(4):359-372.
61. Newcomer JW. Antipsychotic medications: metabolic and cardiovascular risk. *J Clin Psychiatry* 2007; 68 Suppl 4:8-13.
62. Yumru M, Savas HA, Kurt E, Kaya MC, Selek S, Savas E et al. Atypical antipsychotics related metabolic syndrome in bipolar patients. *J Affect Disord* 2007; 98(3):247-252.
63. DuMouchel W, Fram D, Yang X, Mahmoud RA, Grogg AL, Engelhart L et al. Antipsychotics, glycemic disorders, and life-threatening diabetic events: a Bayesian data-mining analysis of the FDA adverse event reporting system (1968-2004). *Ann Clin Psychiatry* 2008; 20(1):21-31.
64. Guo JJ, Keck PE, Jr., Corey-Lisle PK, Li H, Jiang D, Jang R et al. Risk of diabetes mellitus associated with atypical antipsychotic use among patients with bipolar disorder: A retrospective, population-based, case-control study. *J Clin Psychiatry* 2006; 67(7):1055-1061.
65. McKnight RF, Adida M, Budge K, Stockton S, Goodwin GM, Geddes JR. Lithium toxicity profile: a systematic review and meta-analysis. *Lancet* 2012; 379(9817):721-728.
66. Osby U, Brandt L, Correia N, Ekblom A, Soren P. Excess mortality in bipolar and unipolar disorder in Sweden. *Arch Gen Psychiatry* 2001; 58(9):844-850.
67. Tondo L, Isacson G, Baldessarini R. Suicidal behaviour in bipolar disorder: risk and prevention. *CNS Drugs* 2003; 17(7):491-511.
68. Rihmer Z, Kiss K. Bipolar disorders and suicidal behaviour. *Bipolar Disord* 2002; 4 Suppl 1:21-25.
69. Cipriani A, Hawton K, Stockton S, Geddes JR. Lithium in the prevention of suicide in mood disorders: updated systematic review and meta-analysis. *BMJ* 2013; 346:f3646.
70. Caldwell DM, Ades AE, Higgins JP. Simultaneous comparison of multiple treatments: combining direct and indirect evidence. *BMJ* 2005; 331(7521):897-900.
71. Mavridis D, Giannatsi M, Cipriani A, Salanti G. A primer on network meta-analysis with emphasis on mental health. *Evid Based Ment Health* 2015; 18(2):40-46.
72. Lu G, Ades AE. Assessing evidence consistency in mixed treatment comparisons. *J Am Stat Assoc* 2006; 101:447-459.
73. Briggs A, Sculpher M, Claxton K. *Decision Modelling for Health Economic Evaluation*. New York: Oxford University Press; 2006.

74. Fenwick E, Claxton K, Sculpher M. Representing uncertainty: the role of cost-effectiveness acceptability curves. *Health Econ* 2001; 10(8):779-787.
75. Fenwick E, O'Brien BJ, Briggs A. Cost-effectiveness acceptability curves--facts, fallacies and frequently asked questions. *Health Econ* 2004; 13(5):405-415.
76. Desgagne A, Castilloux AM, Angers JF, LeLorier J. The use of the bootstrap statistical method for the pharmacoeconomic cost analysis of skewed data. *PharmacoEconomics* 1998; 13(5 Pt 1):487-497.