Evidence Review Group's Report

Title: Aripiprazole for the treatment and prevention of acute manic and mixed episodes in bipolar disorder in children and adolescents

Produced by School of Health and Related Research (ScHARR), The University of Sheffield

Authors Lesley Uttley, Research Associate , ScHARR, The University of Sheffield, Regent Court, 30 Regent Street, Sheffield, S1 4DA
Ben Kearns, Research Associate, ScHARR, The University of Sheffield, Regent Court, 30 Regent Street, Sheffield, S1 4DA
Matt Stevenson, Professor of Health Technology Assessment, ScHARR, The University of Sheffield, Regent Court, 30 Regent Street, Sheffield, S1 4DA
Ruth Wong, Information Specialist, ScHARR, The University of Sheffield, Regent Court, 30 Regent Street, Sheffield, S1 4DA
Shijie Ren, Research Associate , ScHARR, The University of Sheffield, Regent Court, 30 Regent Street, Sheffield, S1 4DA

Correspondence to Lesley Uttley, Research Associate , ScHARR, The University of Sheffield, Regent Court, 30 Regent Street, Sheffield, S1 4DA

Date completed 5th March 2013

Source of funding: This report was commissioned by the NIHR HTA Programme as project number 09/123/01.

Declared competing interests of the authors

None

Acknowledgements

The authors wish to thank Dr Argyris Stringaris (Consultant Child and Adolescent Psychiatrist, Maudsley Hospital & Senior Lecturer, Institute of Psychiatry); Dr Bernadka Dubicka (Consultant Adolescent Psychiatrist, Lancashire Care Foundation Trust); who provided clinical advice and commented on the draft report. They also wish to thank Dr Isobel Heyman (Consultant Child and Adolescent Psychiatrist, Great Ormond Street Hospital) who reviewed the clinical opinion provided but did not make substantive comments. They also wish to thank Andrea Shippam (Programme Administrator, ScHARR) for her help in preparing and formatting the report.

Rider on responsibility for report

The views expressed in this report are those of the authors and not necessarily those of the NIHR HTA Programme. Any errors are the responsibility of the authors.

This report should be referenced as follows:

Uttley L, Kearns B, Stevenson M, Wong R, Ren S. Aripiprazole for the treatment and prevention of acute manic and mixed episodes in bipolar disorder in children and adolescents. A Single Technology Appraisal. School of Health and Related Research (ScHARR), The University of Sheffield, 2013.

Contributions of authors

Lesley Uttley acted as project lead and systematic reviewer on this assessment, critiqued the manufacturer's definition of the decision problem, led the critique of the clinical effectiveness methods and evidence and contributed to the writing of the report. Ben Kearns and Matt Stevenson critiqued the manufacturer's economic evaluation and contributed to the writing of the report. Ruth Wong critiqued the searches included in the manufacturer's submission and contributed to the writing of the report. Shijie Ren provided statistical advice for this project.

Table of Contents

Tables

Figures

List of abbreviations

ADHD	Attention deficit hyperactive disorder
ADHD-RS-IV	Attention-Deficit Hyperactivity Disorders Rating Scale
AE	Adverse event
ANCOVA	Analysis of covariance
ARI	Aripiprazole
BMI	Body mass index
BMS	Bristol Myers Squibb
CD	Conduct disorder
CDRS-R	Children's Depression Rating Scale-Revised
CGAS	Children's Global Assessment Scale
CGI-BP	Clinical Global Impressions Scale-Bipolar Version
CHMP	Committee for Medicinal Products for Human Use
CI	Confidence interval
CODA	Convergence diagnostic and output analysis
CrI	Credible interval
CSR	Clinical Study Report
EMA	European Medicines Agency
EPAR	European Public Assessment Report
EPS	Extrapyramidal symptoms
EQ-5D	EuroQol 5-Dimension
ERG	Evidence Review Group
FDA	United States Food and Drug Administration
GBI	General Behaviour Inventory Scale
HRQoL	Health related quality of life
ICER	Incremental cost-effectiveness ratio
ITT	Intention to treat
K-SADS-PL	Kiddie-Sads-Present and Lifetime Version
Kg	Kilograms
LOCF	Last observation carried forward
LS	Least squares
MCMC	Markov chain Monte Carlo
Mg	Milligrams
MS	Manufacturers' submission
NHS	National Health Service
NICE	National Institute for Health and Clinical Excellence
OC	Observed cases

ODD	Oppositional defiant disorder
OLA	Olanzapine
PQ-LES-Q	Paediatric Quality of Life Enjoyment and Satisfaction Questionnaire
PSA	Probabilistic sensitivity analysis
PSSRU	Personal Social Services Research Unit
QALY	Quality-adjusted life year
QUE	Quetiapine
RCT	Randomised controlled trial
RIS	Risperidone
SAE	Serious adverse event
SEM	Standard error of mean
SD	Standard deviation
STA	Single Technology Appraisal
TEAE	Treatment emergent adverse event
UK	United Kingdom
WHO	World Health Organisation
YMRS	Young Mania Rating Scale

1 SUMMARY

1.1 Scope of the manufacturers submission

The decision problem addressed in the manufacturers' submission (MS) was based on the Committee for Medicinal Products for Human Use (CHMP) indication for aripiprazole, which is for the treatment of moderate to severe manic episodes in Bipolar I Disorder in adolescents aged 13 years and older. Treatment duration was limited to a period of 12 weeks.

The population outlined in the final scope issued by NICE was children and adolescents with acute manic or mixed episodes associated with bipolar I disorder. The manufacturers presented evidence from trials conducted in the United States which included patients ranging from ages 8 to 17 years. Clinical advisors to the ERG considered the age of the population in the MS to represent a younger population than seen in UK clinical practice for bipolar I disorder. They also raised concerns about the high number of patients with comorbid ADHD in the trial populations included in the MS. This was supported by a statement received by NICE from consultation on behalf of the Royal College of Psychiatrists stating that a high proportion of subjects with comorbid ADHD is likely to reflect a very different population to that seen in clinical practice in the UK. Finally it was noted that the RCTs identified by the manufacturers were likely to be comprised of patients mainly treated as outpatients; this does not match current UK practice where virtually all children and adolescents treated with bipolar 1 disorder would be treated as an inpatient.

The scope issued by NICE described the appropriate comparators as being: i) antipsychotics (olanzapine, quetiapine or risperidone); ii) valproate; iii) lithium; and iv) combination treatment with any of the above. However, the manufacturers justified the exclusion of mood stabilisers such as lithium and valproate on the basis that they are not generally used as monotherapy treatment for children with bipolar disorder and stated that, if used at all, they are used as adjuncts to atypical antipsychotics. The manufacturers concluded that the only relevant comparators are atypical antipsychotics (olanzapine, quetiapine and risperidone). The clinical advisors to the ERG share this view.

The outcomes listed in the MS matched the outcomes outlined in the final scope. However, the MS did not assess the recurrence of manic episodes.

1.2 Summary of clinical effectiveness evidence submitted by the manufacturers

Direct comparison: aripiprazole versus placebo

The clinical effectiveness evidence in the MS for aripiprazole versus placebo was based predominantly on the following study:

- A phase III randomised controlled trial (RCT) conducted in 296 children and adolescents aged 10-17 years (NCT00110461). The study was double blind and placebo controlled and consisted of an acute 4 week phase and a 26 week maintenance phase. The duration of the study was therefore longer than the recommended 12 week CHMP indication.

The manufacturers also presented a meta-analysis with trial NCT00110461 and a smaller trial of 43 patients which also examined aripiprazole as per the decision problem but included only patients with comorbid ADHD and also included patients with bipolar II disorder. This trial was:

- An RCT conducted in children and adolescents aged 8-17 years (NCT00116259)

The main findings of this meta-analysis did not significantly alter the results of trial NCT00110461 alone.

The phase III trial NCT00110461 was considered by the ERG to be relevant to the decision problem as specified in the scope, but concurred with the manufacturers that the NCT00116259[1] trial should not be included in the base case due to the different patient population.

The searches for clinical evidence in the systematic reviews presented in the MS were limited to January 2012 which was one year prior to the MS. Additionally the manufacturers' search strategy was not adequate to capture non-RCT evidence for adverse events. Following a request from the ERG, the manufacturers used non-systematic approaches to update the searches to January 2013. The ERG performed updates of the systematic searches up to January 2013 and requested clarification from the manufacturers regarding four completed clinical trial records which appeared relevant to the decision problem. The manufacturers clarified the exclusion of three of the trials but stated that the fourth trial had not been identified. However they did provide a brief synopsis of this single arm study of aripiprazole monotherapy in children and adolescents aged 7 to 18 years old, diagnosed with bipolar I disorder, manic or mixed episode. No additional phase III randomised controlled trials were identified by the ERG or clinical advisors to the ERG.

Clinical efficacy: The results of the phase III trial NCT00110461 showed statistically significant improvement compared with placebo in the primary efficacy endpoint which was the mean change from baseline to week 4 in the total Young Mania Rating Score (YMRS).

Significant improvements were also documented for the secondary efficacy endpoints: Children's Global Assessment Scale (CGAS); Clinical Global Impressions Scale-Bipolar Version (CGI-BP) Severity Score for mania and for overall bipolar illness; General Behaviour Inventory Scale (GBI) Total Score for mania and for the Attention Deficit Hyperactive Disorder ADHD Rating Scale (ADHD-RS-IV).

Both the aripiprazole 10 mg and 30 mg arms had significantly higher percentages of responders (defined as a ≥50% reduction in YMRS score) compared with the placebo arm at week 4 using the Last Observation Carried Forward (LOCF) data set.

Safety profile: Data from the phase III trial NCT00110461 using the pooled data from the 10mg and 30mg aripiprazole treatment arms versus placebo indicated the following safety profile versus placebo (at 4 weeks):

- Aripiprazole was significantly more likely to cause extrapyramidal symptoms (EPS) than placebo ($p<0.001$);
- Aripiprazole was significantly more likely to cause somnolence than placebo ($p<0.001$);
- There were no significant differences between aripiprazole and placebo for clinically significant increases in weight gain or clinically significant increases in prolactin.

At the end of the extension phase (week 30) [redacted]

Indirect comparison: aripiprazole versus olanzapine; risperidone; and quetiapine

Data from the phase III trial NCT00110461 of aripiprazole; a study of risperidone (Haas 2009); a study of quetiapine (Study 149); and a study of olanzapine (Tohen 2007) were used to perform an indirect comparison. All studies compared antipsychotic treatment to no treatment (placebo) at 3 weeks. Where there were more than one treatment dose, the data from the multiple treatment arms were pooled. The indirect comparison was performed using a network meta-analysis. Two further studies of risperidone were identified for inclusion in

the meta-analysis (Pavuluri 2010 and Geller 2012) but were excluded from the main analysis due to increasing uncertainty in the meta-analysis. The ERG noted that the trial population in the Geller 2012 study was markedly different to the other included trials however did not consider the exclusion of the Pavuluri 2010 study to be valid. Both studies, along with the smaller aripiprazole trial NCT00116259 were included in sensitivity analyses by the manufacturers.

Clinical efficacy: Using the YMRS response, aripiprazole performed similarly to the comparators olanzapine, risperidone and quetiapine.

Safety profile: Data from the phase III trial NCT00110461; Haas 2009; Study 149; Tohen 2007 (pooled doses where there were multiple treatment arm doses) were compared for adverse events of EPS; clinically significant weight gain; clinically significant increase in prolactin and somnolence.

- For clinically significant weight gain aripiprazole performed significantly better than olanzapine (median RR vs. aripiprazole: 12.52 [95% CrI 2.31-76.22]) and quetiapine (median RR vs. aripiprazole: 11.1 [95% CrI 1.30-116.1]) but not risperidone (median 1.19; 95% CrI 0.22-6.94).
- For clinically significant prolactin increase aripiprazole performed significantly better than olanzapine (median RR vs. aripiprazole: 175.70 [95% CrI 10.86-6414]); risperidone (median RR vs. aripiprazole: 139.80 [95% CrI 5.52-7202]) or quetiapine (median RR vs. aripiprazole: 31.22 [95% CrI 1.81-1191]).
- For the adverse events of EPS and somnolence, aripiprazole performed less favourably than risperidone and quetiapine, but not significantly so. No data for olanzapine on somnolence were available for this comparison.

1.3 Summary of the ERG's critique of clinical effectiveness evidence submitted

The ERG considers the US paediatric bipolar I population included in trial NCT00110461to be discrepant to UK population according to the low mean age; and high prevalence of comorbid ADHD. Additionally the severity of the patients included in the trial population in the MS is unlikely to reflect clinical practice in the UK. This is due to the inclusion criteria employed in trial NCT00110461 stipulating that suicidal patients were excluded from participating in the study. The manufacturers were unable to provide the ERG with data on the number of trial patients who were inpatients (as would be the case in UK clinical practice) which also suggests that the population in the MS may not reflect the UK paediatric bipolar I population.

On the basis of the evidence from RCTs included in the MS, aripiprazole has a similar efficacy profile, in terms of YMRS reduction, as olanzapine, risperidone and quetiapine. There is no clear evidence that aripiprazole has a worse side effect profile than olanzapine, risperidone and quetiapine.

As the trial NCT00110461 duration was 30 weeks, the duration of maintenance of effect of only 12 weeks of aripiprazole treatment is unknown. No recurrence data were provided by the manufacturers to indicate how long patients in the included trial remain stable following discontinuation of antipsychotic treatment. The focus of the MS was treatment of the acute phase. The use of aripiprazole as maintenance therapy, as may be used in clinical practice, is outside the CHMP's recommended duration of treatment. However, the 30 week data indicate that the safety profile of aripiprazole during the extension phase was acceptable.

Pooling doses from treatment arms with multiple doses may not necessarily be appropriate, as it is possible that different doses are associated with different efficacies and side effects. The ERG asked that the network meta-analyses be performed having separated the different doses, however the manufacturers responded that this was not possible to run these additional analyses within the time permitted to respond to clarifications.

The ERG requested the manufacturers to undertake network meta-analyses using a random effects model, rather than a fixed effects model, as it was likely that there was heterogeneity within the RCTs. The manufacturers did not undertake these analyses, which were performed by the ERG.

It is noted that not all the information requested by the ERG were made available. It is unclear whether if these data were known whether this would have an impact on the clinical interpretation. Data on adherence was collected but not provided in the MS. The categories for which incomplete information was provided included: comorbid ADHD; the numbers in age subgroups; rapid cyclers; mixed/ manic episode; numbers in receipt of psychotherapy; and the numbers of patients in community versus inpatient care.

1.4 Summary of cost effectiveness submitted evidence by the manufacturer

The manufacturers supplied a *de novo* cohort Markov model constructed in Microsoft Excel©. A sequence of up-to four treatment lines were modelled, of which the first three related to treatment with an antipsychotic drug and had the same structure: the acute (three-week inpatient treatment) phase; sub-acute (inpatient treatment of responders) phase; and the maintenance phase (outpatient treatment followed by withdrawal of treatment). The fourth

treatment line was of lithium for therapy resistant patients; both inpatient and outpatient treatment was modelled. Patients moved down treatment lines if they discontinued drug use or failed to respond in the acute phase, or if they relapsed during the sub-acute phase. Patients could also die at any point.

Data relating to the effectiveness and safety profile (incidence of weight gain, somnolence and EPS) for each drug were taken from the manufacturers' network meta-analysis.

For inpatients, resource use was defined by the hospital structure, with costs taken from NHS reference costs 2010/11 (code MHIPC1; NHS Trusts Mental Health Inpatients – Children). The manufacturers assumed that this included costs relating to adverse events, but not the cost of the antipsychotic. Out-of-hospital resource use was based on expert opinion, with costs taken from the Personal Social Services Research Unit. Drug costs (where appropriate) were included separately, as were costs related to weight-gain.

The manufacturers were not able to identify any preference-based measures of health-related quality of life for paediatric bipolar disorder. Nor were they able to identify any reliable methods for mapping to these. Instead the manufacturers used EQ-5D data from an adult UK population with bipolar disorder. These data were modelled as multiplicative weights applied to general population EQ-5D values. The utility weight for weight gain was taken from the general population, weights for somnolence and EPS came from patients with schizophrenia.

For their base-case analysis the manufacturers used a comparator treatment sequence of risperidone, quetiapine, olanzapine. The use of aripiprazole instead of olanzapine and at any point in the treatment sequence was considered, resulting in three intervention sequences (quetiapine was always after risperidone). Treatment with each antipsychotic is set to an average of 10 to 12 weeks, reflecting the CHMP opinion which restricted the use of aripiprazole (the only intervention with a licence in children and adolescents) to 12 weeks.
Cost-effectiveness results were similar for the four treatment sequences. However, the use of aripiprazole as a second line treatment following risperidone resulted in both the lowest total costs (£74,133) and the highest total QALYs (2.525). The strategy where aripiprazole was not included had both the highest total costs (£75,066) and the lowest total QALYs (2.516).

1.5 Summary of the ERG's critique of cost effectiveness evidence submitted

In general the ERG is satisfied that the economic evidence submitted by the manufacturers does not represent a biased assessment of the cost-effectiveness of aripiprazole. Minor

changes were made to include a half-cycle correction; adjust the discounting formula used; amend the mortality calculations; and impose a logical constraint on the PSA inputs. These had a negligible impact on the results. However, the ERG considers that there are three topics that were not addressed in sufficient detail in the manufacturers' submission. These are the modelled length of treatment for aripiprazole, the impact of personalising treatment sequences to reflect the patients' needs and the omission of strategies assessing sequences using four antipsychotic interventions where necessary. These are discussed in turn.

The ERG notes that within the manufacturers' model it is possible for patients to remain on aripiprazole treatment for longer than twelve weeks. In addition, the ERG has been advised by clinical experts that the length of antipsychotic treatment is typically closer to twelve months than twelve weeks. However, the ERG also notes that the use of aripiprazole has only received CHMP approval for a maximum of twelve weeks. Because of this, two different treatment durations could be modelled: one reflecting real-world prescribing with an average duration of twelve months, the other reflecting the licenced duration for aripiprazole of a maximum of twelve weeks. The manufacturers' model, which sets treatment duration to an average of ten to twelve weeks, strictly models neither, although was more representative of the licenced duration.

The ERG amended the manufacturers' model to have a maximum treatment duration (for all antipsychotics) of twelve weeks. The manufacturers (after a request from the ERG) provided an amended version of their model to have an average of twelve months of antipsychotic treatment. Both total costs and total QALYs showed a reduction in both of the two new models, but the substantive conclusions of the manufacturers' base-case analysis remained unchanged.

The manufacturers did not present any of their results as incremental analyses. The ERG recalculated the results as incremental analyses. This showed that S2 dominated all of the other treatment strategies in the manufacturers' base-case. This domination was found to hold for nearly all of the one-way sensitivity analyses and scenario analyses conducted by the manufacturer, and may imply that aripiprazole should be recommended as second-line treatment. However, both the clinical advisors to the ERG and those to the manufacturers stressed the importance of tailoring the treatment sequence to reflect an individual's needs (based on factors such as severity of symptoms; side-effect profile; or comorbidities for example). There are limited data available to model treatment within sub-groups, so the ERG conducted an exploratory scenario analysis to look at the possible implications of personalised

medicine. The results showed that only small changes in the modelled results – typically no more than 2% of either the total costs or total QALYs – for each treatment sequence were needed for that strategy to become cost-effective (assuming a willingness to pay of £20,000 or £30,000 per QALY). These results suggest that the actual place of aripiprazole within a treatment sequence is likely to depend on individual circumstances.

The clinical advisors to the ERG indicated that if a patient had not responded to three antipsychotic interventions that they would use the remainder antipsychotic rather than declaring the patient treatment resistant. This was not possible to evaluate within the manufacturers model, although there is no evidence that this would substantially alter the conclusions.

1.6 ERG commentary on the robustness of evidence submitted by the manufacturers

1.6.1 Strengths

The ERG identified a number of strengths in terms of the robustness of evidence in the submission, including the following points.

- The decision problem addressed in the MS was relevant to the NICE scope
- Relevant evidence in terms of placebo controlled trial were used for the indirect comparison with other placebo-controlled trials
- The pivotal phase III trial comparing aripiprazole with placebo was of reasonable methodological quality, and measured a range of outcomes that are relevant to the decision problem.
- The studies and outcomes included in the indirect comparison were appropriate
- The economic model used appears to be robust and transparent, allowing for the analysis of uncertainty in the model inputs.

1.6.2 Weaknesses and areas of uncertainty

With respect to the clinical effectiveness evidence the key areas of uncertainty identified by the ERG are as follows:

- The trial population in the MS are likely to be discrepant to the UK clinical population
- Incomplete information were available on a selection of clinical parameters
- Caution should be applied when interpreting results presented in the MS using pooled intervention doses from multiple treatment arms

With respect to the economic evaluation the key areas of uncertainty identified by the ERG are:

- A treatment sequence incorporating all four antipsychotics was not included.
- There is a lack of evidence to show how different sub-groups would respond to different treatment sequences. Clinical advisors to the ERG and clinical advisors to the manufacturers both believe that this is important. Exploratory analyses indicated that small differences could have a large impact on the cost-effectiveness results.

1.7 Summary of additional work undertaken by the ERG

The ERG repeated and updated the searches conducted by the manufacturers to January 2012, up to January 2013. Additionally the ERG carried out supplementary searches for non-RCT evidence. All uniquely identified records from both additional searches were retrieved and reviewed by the ERG. Records identified as being potentially relevant to the decision problem are discussed in this report.

The ERG undertook network meta-analyses using a random effects model and also using a Bayesian prior where observed values were zero. The central estimates of efficacy were broadly similar to those produced by the manufacturers, although as would be expected the uncertainty around these point estimates were increased.

The ERG redisplayed the manufacturers' original results as incremental analyses; these should that for the majority of analyses aripiprazole second-line dominated all of the other treatment sequences. The ERG performed additional analyses of the manufacturers' PSA results and conducted an exploratory analysis into the impact of personalised medicine. This showed that there was great uncertainty in the results: only small changes were required for each treatment strategy to become a viable alternative to second-line aripiprazole.

The ERG explored the potential implications of two different uses for aripiprazole: one use reflected its licenced duration of a maximum of twelve weeks; the other reflected its real-life use of an average of twelve months. Cost-effectiveness results for these two situations did not show a noticeable difference from the manufacturers' base-case results.

2 BACKGROUND

2.1 Critique of manufacturers' description of underlying health problem

The manufacturers' (Otsuka Pharmaceutical Co. and Bristol Myers Squibb) description of the health problem was based on information from the National Institute of Mental Health[2]; NICE clinical guidelines[3] and a published study.[4] Bipolar disorder is described as a disease in which a patient's mood and energy levels can oscillate, affecting their ability to perform everyday tasks such as attending school or socialising with peers. Bipolar I disorder is characterised by at least one manic episode, with periods of major depression. If manic and depressive phases overlap and a patient experiences manic and depressive symptoms simultaneously or in close succession, this is defined as a mixed state. The manufacturers' submission (MS) describes the main symptoms of manic episodes in bipolar I disorder in children and adolescents to be:

- Poor concentration
- Little need for sleep
- Poor temper control
- Reckless behaviour and lack of self-control
- Euphoria/very elevated mood
- Grandiosity
- Irritability
- Psychosis (loss of contact with reality)ToMyShow1

The MS states that compared with adult-onset bipolar disorder, children and adolescents often experience more severe manifestations, which may lead to worse outcomes in the long term.[5] Moreover, the quality of life experienced by children and adolescents with bipolar disorder is severely reduced[6], particularly with respect to psychosocial dimensions of health including social and family well-being.[7] The Evidence Review Group (ERG) and the clinical advisors to the ERG considers the manufacturers' description of the underlying health problem to be accurate.

The manufacturers' description of clinical prevalence was based on a published study by Soutullo et al., 2005[8] and a study report by the Royal College of Psychiatrists in 2001[9] of child and adolescent inpatients in the United Kingdom. The manufacturers estimated the prevalence of bipolar I disorder to be 136 patients hospitalised per year on page 29 of the MS. The manufacturers acknowledge that the figure of 136 is likely to have risen, as the estimate is based on data from 1996. Two of the clinical advisors to the ERG stated that the prevalence figure in the MS was insufficiently referenced and that more recent UK studies should be

cited. For example a more recent epidemiological study of bipolar disorder was conducted in 2009 by Stringaris *et al.*,[10] The author of this paper, who is one of the clinical advisors to the ERG, indicated that they would expect the figure to be up to 250; a higher figure than suggested by the manufacturer. However, the clinical advisors to the ERG were in broad agreement that the number of cases were low.

The manufacturers presented two summary tables of prevalence studies from thirteen countries in an appendix to the MS (pages 264/5) in which the prevalence ranges from 0% to 7.2%. These values are likely to reflect a wide variation between countries in the diagnosis of bipolar disorder and its subtypes.[11] The World Health Organisation (WHO) estimates the total lifetime prevalence of bipolar I disorder to be 0.6% (Merikangas *et al.,* 2012[12]). However, this figure cannot simply be extrapolated to children and adolescents since it is reported that the increase in the number of diagnoses of bipolar disorder is greater in children and adolescents than in adults (Moreno *et al.,* 2007[13]). The ERG acknowledge that there is considerable difficulty in estimating a figure of paediatric bipolar I disorder in the UK for the following reasons:

i. the estimated figure above relates to inpatients only. The number of paediatric bipolar I disorder patients in community practice is unknown.
ii. overlapping diagnostic criteria of conditions such as attention deficit hyperactivity disorder (ADHD) (Youngstrom *et al.,* 2009[14])
iii. underdiagnoses among UK clinicians due more restrictive criteria regarding irritability in the NICE guidelines as opposed to the DSM-IV criteria[15,10] and reluctance to prematurely diagnose young people with a life-long mental illness such a bipolar disorder.[16]

In view of the difficulties discussed above, the ERG considers that the manufacturers' estimation of clinical prevalence is not inappropriate to the decision problem under consideration but that a more recent and accurate estimate of clinical prevalence is lacking.

2.2 Critique of manufacturers' overview of current service provision

The MS referred to the NICE Clinical Guideline 38[3] which states that the only drug licensed for use in children and adolescents with bipolar disorder is lithium, which should only be used as an adjunct to atypical antipsychotics. The MS also stated that the Royal College of Paediatrics and Child Health[17] recommend that unlicensed medications may be prescribed where there are no suitable alternatives and where the use is accepted by a professional

opinion. The clinical advisors to the ERG agree that atypical antipsychotics: aripiprazole; olanzapine; quetiapine; and risperidone, are all used as first line treatment for acute manic and mixed episodes in paediatric bipolar I disorder and that the efficacy of antipsychotics in acute mania has been documented in a meta-analysis of pharmacological drugs.[18] As of 13th December 2012, aripiprazole has received a positive opinion from the Committee for Medicinal Products for Human Use (CHMP). Therefore whilst other atypical antipsychotics: olanzapine; risperidone; and quetiapine are commonly used by paediatric psychiatrists in the treatment of acute manic and mixed episodes in bipolar I disorder, only aripiprazole has received a positive CHMP opinion for use in adolescents over the age of 13 years. The MS also stated that [redacted]

The MS estimated the length of hospitalisation for paediatric patients experiencing an acute manic episode based on the opinion of their clinical experts who stated that patients would be hospitalised and that they would remain hospitalised for up to two months under observation. One of the clinical advisors to the ERG said that they would expect that most patients would be in hospital for at least 14 weeks. Another of the clinical advisors to the ERG noted that depending on the psychiatric health service available, patients may be transferred to "day patient" status after 3 or 4 weeks if patients have responded to medication. However, both clinical advisors agreed that it is not uncommon for patients to be in hospital for 14 weeks.

The manufacturers proposed a treatment pathway based on the responses of three clinical advisors comprising of risperidone as first-line use, quetiapine as second line and olanzapine as third line. The clinical advisors to the ERG did not agree that they have a preferred order of treatment for prescribing atypical antipsychotics but said that the clinical presentation of the patient will determine which drug is tried first. For example, one clinical advisor stated that if a patient is agitated/irritated then olanzapine may be used first as it has a sedative effect or if a patient is depressed then quetiapine may be considered before the other atypical antipsychotics. This clinical advisor stated that aripiprazole is rarely used first line as a greater degree of sedation is often required. The second clinical advisor agreed with these statements and added that aripiprazole is used in order to avoid weight gain and increased prolactin levels but that, in line with the adult data[19,20] if a patient was agitated or depressed then other treatments may be preferable. The third clinical advisor reiterated that whilst aripiprazole seems a safe option in both acute control and maintenance, usually a more sedative drug is required in acute control. This advisor added that this opinion is based on clinical experience rather than trial evidence.

3 CRITIQUE OF MANUFACTURERS' DEFINITION OF DECISION PROBLEM

A summary of the decision problem (Table 1) as outlined in the final scope issued by NICE which was defined in the context of NICE Clinical Guideline No. 38[3] and addressed in the manufacturers' submission is presented in Table 1.

Table 1: Decision problem as outlined in the final scope issued by NICE and addressed in the manufacturers' submission (based on pages 38-40 of MS but amended by ERG to reflect their opinion of the submission)

	Decision problem outlined in final scope issued by NICE	Decision problem addressed in the submission
Population	Children and adolescents with acute manic or mixed episodes associated with bipolar I disorder	Adolescents with acute manic episodes associated with bipolar I disorder
Intervention	Aripiprazole for the treatment and prevention of acute manic and mixed episodes in bipolar I disorder in children and adolescents	Aripiprazole for the treatment of adolescents with bipolar I disorder mania
Comparator(s)	• Antipsychotics (e.g., olanzapine, quetiapine or risperidone) • valproate • lithium • combination treatment with any of the above	• Atypical antipsychotics (olanzapine, quetiapine or risperidone) • combination treatment of any atypical antipsychotic with either valproate or lithium
Outcomes	• response rate • range and severity of symptoms of mania and depression • recurrence of manic episodes • body mass index (adjusted for the child's age and gender). • adverse effects of treatment • health-related quality of life.	• response rate • range and severity of symptoms of mania and depression • recurrence of manic episodes • body mass index • adverse effects of treatment • health-related quality of life.

Economic analysis	The reference case stipulates that the cost effectiveness of treatments should be expressed in terms of incremental cost per quality-adjusted life year (QALY). The reference case stipulates that the time horizon for estimating clinical and cost effectiveness should be sufficiently long to reflect any differences in costs or outcomes between the technologies being compared. Costs will be considered from an NHS and Personal Social Services perspective.	As final scope. The time horizon for modelling is 3 years whereby patients enter the model at aged 15 until adulthood. The time horizon does not include treatment of bipolar disorder into adulthood, as this is a different indication and is appraised separately by NICE.
Subgroups to be considered	If evidence allows, the effectiveness of aripiprazole in pre-pubescent children compared with post-pubescent children will be assessed. If evidence allows the effectiveness of aripiprazole alone or in combination with lithium or valproate will be assessed.	Subgroup analyses of efficacy by: • Age group 10-12 years and 13-17 years • comorbid ADHD
Special considerations, including issues related to equity or equality	Guidance will only be issued in accordance with the marketing authorisation.	The atypical antipsychotics risperidone, quetiapine and olanzapine are used as comparators in the economic model (all of which are currently used off-label for the acute or maintenance treatment of children or adolescents with bipolar I disorder). The original model also allowed off-label use of aripiprazole beyond 12 weeks.

3.1 Population

The population described in the final scope issued by NICE, issued in accordance with NICE Clinical Guideline no. 38[3] was children and adolescents with acute manic or mixed episodes associated with bipolar I disorder.

Aripiprazole received a positive CHMP opinion on 14th December 2012 for the treatment (of up to 12 weeks' duration) of moderate to severe manic episodes in Bipolar I Disorder in adolescents aged 13 years and older. The MS states (page 19) that the CHMP concluded that the available data to date raised safety concerns mostly regarding weight gain and EPS symptoms, especially in the young bipolar I disorder population aged 10-12 years. The CHMP therefore noted that the safety profile was not favourable for the younger population

(10-12 years) and therefore concluded that the benefit –risk balance was only positive in the paediatric bipolar I disorder population aged 13 years and older.

The population included in the MS for the assessment of clinical efficacy are aged 8-17 years as per the inclusion criteria of the included trials NCT00110461[21] and NCT00116259.[1] Clinical advisors to the ERG stated that it is very rare to see a child as young as 10 years old in the UK who is diagnosed with bipolar disorder and that the peak age onset would be later than 10 years, as indicated in the MS. Trial NCT00110461 included patients described as having a confirmed DSM-IV diagnosis of bipolar I disorder whilst trial NCT00116259[1] included patients with bipolar II disorder as well as bipolar I disorder. Both trials describe patients as currently in manic or mixed states. It was unclear from the MS what proportion of included patients from trial NCT00110461[21] were hospital inpatients and therefore the ERG requested clarification on this issue. Clinical advisors to the ERG have indicated that in the UK it would be extremely rare for patients with moderate to severe mania to be managed in the community and that such patients would be almost certainly be inpatients. The manufacturers responded that the data on the number of inpatients/outpatients were not reported in the clinical study report (CSR) and are therefore not provided. The evidence presented for the effectiveness of aripiprazole throughout the MS is limited to trials of children and adolescents under 18 years of age. However a substantial number of trials in bipolar I disorder have been conducted in the adult population. These data are widely available and form the basis of the FDA and EMA licence for use of aripiprazole in bipolar I disorder.

Clinical advisors to the ERG have suggested that, on the basis of the low age range of trials included in the MS, it may have been more appropriate to use published evidence from the adult population than data from children who are younger than the typical clinical profile of paediatric bipolar I disorder patients in the UK. This may be particularly relevant in light of the CHMP restriction of aripiprazole to patients who are 13-17 years of age. The ERG requested clarification from the manufacturers on whether the adult data for aripiprazole would have been relevant to this assessment. The manufacturers responded that, from the expert opinion they had received, children and adolescents have different symptoms to adults and are treated more intensively. These symptoms include more severe manifestations; more rapid changes in disease states; and being more prone to some adverse events such as weight gain.[22-24] The manufacturers also responded that the indication in children is restricted to the treatment of acute episodes, rather than long term maintenance as for adults. Additionally, it was stated that it would not be appropriate to generalise adult clinical data to children and

adolescents given the particular clinical characteristics of bipolar disorder in younger individuals, and the mismatch in treatment duration (page 1 of the clarification response). One clinical advisor agreed with the manufacturers that adult data is only extrapolated to children and adolescents in clinical practice with caution due to the paucity of data in young people; but that adult data cannot be regarded as a valid evidence base for children and adolescents. Two of the clinical advisors to the ERG stated that failure to incorporate evidence from adult data is a missed opportunity, particularly in areas such as quality of life, since the transition from childhood to adulthood at the age of 18 is not a sudden change. However, as the final scope issued by NICE restricted the inclusion of evidence to patients under 18 years of age the ERG considers the age of the population included by the manufacturers in the MS to be justified.

In trial NCT00110461[21] (page 139 of MS), 153/196 (51.7%) patients are reported to have current comorbid ADHD whilst in trial NCT00116259[1] all patients had comorbid ADHD. Clinical advisors to the ERG have indicated that these high numbers of patients with comorbid ADHD are not likely to reflect the phenotype of patients diagnosed with bipolar I disorder in the UK and may indicate complex ADHD rather than bipolar disorder.

This is supported by a statement received by NICE from consultation for this assessment from Dr David Coghill (on behalf of the Royal College of Psychiatrists) who stated that "any trial that includes individuals with bipolar 2 or NOS or a high proportion of subjects with comorbid ADHD is likely to reflect a very different population to that seen in clinical practice in the UK".

It is the opinion of the ERG that caution should be applied when interpreting the evidence from the trials included in the MS of US paediatric patients with bipolar I disorder. The manufacturers acknowledge (page 141 of the MS) the potential limitation of external validity due to differences in the diagnosis of bipolar disorder between the US and the UK. It is therefore unclear to what extent the populations included in the MS reflect the relevant paediatric bipolar I disorder population in UK clinical practice due to the low age range, high prevalence of comorbid ADHD and high numbers managed in community care.

3.2 Intervention

The intervention described in the final scope issued by NICE was aripiprazole for the treatment and prevention of acute manic and mixed episodes in bipolar I disorder in children and adolescents. Due to the CHMP positive opinion's restriction of treatment (of moderate to

severe manic episodes) beyond 12 weeks' duration, the intervention described in the MS is aripiprazole for the treatment of adolescents with bipolar I disorder mania.

Aripiprazole (UK brand name: Abilify®) is an atypical antipsychotic with partial dopamine D2 and D3 agonistic properties. The mechanism of action of aripiprazole differs from other atypical antipsychotics because it acts as a D2 partial agonist rather than antagonizing the D2 receptor.

Aripiprazole is available as tablets, orodispersible tablets or oral solution for the treatment of children and adolescents. The acquisition cost of 28 tablets or orodispersible tablets (10mg or 15mg) is £95.74. The acquisition cost of 28 tablets or orodispersible tablets (30mg) is £191.47. For patients who respond to aripiprazole, the expected length of a course of treatment is 12 weeks. For a course of 12 weeks (84 days), the 10 mg dose would cost £287.22. This would be the same for a 15 mg dose. A course of 30 mg dose would cost £574.41. A 12-week course of aripiprazole may be considered as a suitable duration for the treatment of a manic or mixed episode but the current license would not cover the prescription of aripiprazole as maintenance therapy to prevent further episodes.

The expected recommended dose of aripiprazole is 10 mg/day on a once-a-day schedule. The CHMP recommend (page 19 of the MS) that an increase over 10 mg should only be performed under strict surveillance. Treatment should be initiated at 2 mg (using oral solution 1 mg/ml) for 2 days, titrated to 5 mg for 2 additional days to reach the recommended daily dose of 10 mg. Where appropriate, subsequent dose increases can be administered in 5 mg increments without exceeding the maximum daily dose of 30 mg.

The manufacturers stated (MS pages 22/23) that there is uncertainty around the average number of courses of treatment required and that due to the frequent relapsing nature of paediatric bipolar I disorder, many patients would be expected to require subsequent courses of treatment after the first course. As per treatment with other atypical antipsychotics, mood stabilisers such as lithium and valproate can be added to aripiprazole if symptoms persist.

Lithium is currently the only medication with a marketing authorisation for the treatment of manic episodes associated with bipolar I disorder in children and adolescents. Valproate should normally be avoided in girls and young women because of teratogenicity (foetal malformation) risks during pregnancy and risk of polycystic ovary syndrome. Sedation

medication (benzodiazepine) can also be added if necessary. Anticholinergic therapy is used to treat extrapyramidal symptoms (EPS).

General management principles for the drug treatment of bipolar disorder in children and adolescents include starting at lower doses than in adults and close monitoring, as this population may be more susceptible to adverse events, such as sedation, obesity, extrapyramidal symptoms, metabolic changes and raised prolactin.[3] NICE guidelines[3] state that children and adolescents experiencing acute mania should be treated according to the recommendations for adults with bipolar disorder, with the exception that therapy should be initiated at lower doses. The following factors should also be considered:

i. Height and weight should be checked and monitored regularly.
ii. Prolactin levels should be monitored.
iii. That there is a risk of increased prolactin levels with risperidone and of weight gain with olanzapine.

Aripiprazole is also available as a solution for intramuscular injection but the intravenous formulation of aripiprazole is not under consideration in this assessment as it was not included in the final scope issued by NICE.

3.3 Comparators

The scope issued by NICE issued in accordance with NICE Clinical Guideline No. 38[3] described the appropriate comparators to be i) antipsychotics (olanzapine, quetiapine or risperidone); ii) valproate; iii) lithium; and iv) combination treatment with any of the above. However, the manufacturers stated that (MS page 38) mood stabilisers such as lithium and valproate are not generally used as monotherapy treatment for children with bipolar disorder and that clinical opinion is that, if used at all, they are used as adjuncts to atypical antipsychotics. The manufacturers conclude that the only relevant comparators are atypical antipsychotics (olanzapine, quetiapine and risperidone). The clinical advisors to the ERG share this view.

The pivotal trials (NCT00110461[21] and NCT00116259[1]) presented by the manufacturers for the demonstration of the efficacy and safety of aripiprazole use placebo as the comparator. The guidance produced by the EMA[25] in 2001 for clinical investigation of bipolar disorder recommends that efficacy should be studied using trials with active and placebo controls. Whilst this guidance was not followed in the included trials, which were both conducted

between 2005- 2008, it is noted that risperidone; olanzapine and quetiapine are being used "off-label" as they have not received a license for use in children and adolescents The manufacturers compare the effectiveness of aripiprazole versus the atypical antipsychotics using a network meta-analysis.

3.4 Outcomes

The outcomes listed in the statement of the decision problem (MS page 39) matched the outcomes outlined in the final scope.

The primary outcome for response rate was defined as the change from baseline to week 4 on the Young Mania Rating Scale (YMRS) total score. Response rate was also measured ('response' being defined as ≥50% reduction from YMRS total score).

Secondary outcomes were stated in the decision problem to be: i) range and severity of symptoms of mania and depression; ii) recurrence of manic episodes; iii) body mass index; iv) adverse effects of treatment; and v) health-related quality of life.

Recurrence of manic episodes was not reported in the MS despite being listed in the decision problem. The EMA defines recurrence as "a re-emergence of symptoms (new episode) after a time with no or minimal symptoms" whilst relapse is defined as "an increase in symptomatology immediately or almost immediately after a time with no or minimal symptoms".[25] In addition [redacted] however the MS does report (MS page 80) [redacted] Body mass index (BMI) and adverse effects of treatment are reported in the MS for both included trials. Health-related quality of life was measured using the Paediatric Quality of Life Enjoyment and Satisfaction Questionnaire (PQ-LES-Q) in trial NCT00110461. Quality of life was not measured in trial NCT00116259.[1]

3.5 Other relevant factors

The MS contained a Section on equality issues (MS page 37). The manufacturers stated that there were no equality issues relating to the use of aripiprazole under its licence.

The updated European Public Assessment Report (EPAR) for Abilify® was published on 20/03/2012[26] by the European Medicines Agency.

The US patent for aripiprazole which belongs to Otsuka expires on October 20, 2014. However, it is reported in the Orange Book on the FDA website[27] that there is a paediatric extension, and therefore a generic may not become available until at least April 20, 2015.

4. CLINICAL EFFECTIVENESS

4.1 Systematic reviews for clinical efficacy

Two systematic reviews were performed for the assessment of clinical effectiveness in the MS. The objective of the first systematic review was to identify all relevant clinical information available for the treatment and prevention of acute manic and mixed episodes in bipolar disorder in children and adolescents with aripiprazole (MS page 43). The review was based on the search and inclusion strategy of a previous systematic review (commissioned by NICE in 2005[28]) which did not identify any relevant RCT or non-RCT evidence for aripiprazole in children and adolescents with bipolar I disorder (MS page 43).

The second systematic review was designed with the objective of identifying all relevant clinical information available for the treatment and prevention of acute manic and mixed episodes in bipolar disorder in children and adolescents with the following comparators (MS page 92):

- Risperidone
- Quetiapine
- Olanzapine
- Combination of any of the above with lithium or valproate

This review was also based on the same search and inclusion strategy as the NICE 2005 review[28], and identified three randomised controlled trials that examined the use of antipsychotics in the treatment of mania in children and adolescents. However, only the results from one trial (DelBello 2002[29]) were included in this update. Of the two other excluded studies, one was an open label trial including some bipolar II patients (Pavuluri 2004[30]), and one was a semi-randomised controlled trial which also included bipolar II patients (Biederman 2005[31]). For the long term management of children and adolescents with bipolar disorder, the only study identified by the previous systematic review was excluded not only because it examined lithium, but also because it included a mixture of bipolar I and II patients (Findling 2005[32]).

It was not clear from the MS why the searches for both reviews were limited up to January 2012 which makes the evidence generated from their searches one year out of date and therefore clarification was requested from the manufacturers. The manufacturers did update the searches on the 4th February 2013 in response to the clarification questions raised by the ERG. The manufacturers stated that rather than repeating their searches a non-systematic approach was used to find further studies since January 2012 due to time constraints. Only

PubMed, Google Scholar and ClinicalTrials.gov were searched by the manufacturers but Embase and Cochrane Library were not searched. Four RCTs and three non-RCTs were identified (page 1 of clarification response).

The ERG repeated and updated the searches until January 2013 using the systematic approach in the manufacturer's submission. Database searches were repeated and updated by using strategies provided in the MS. Searches of congress websites as listed in Appendix 1b (MS page 267) were not reproduced by the ERG as it was not clear which terms were used in these websites. The ERG had access to the databases used by the manufacturers with the correct host interfaces. The manufacturers reported that 4904 records were found. Repeat and update searches by the ERG have found a total of 5277 records, of which, 425 were published in 2012, which represents a significant number of potentially relevant records that were missed by the manufacturers searches that were conducted up to January 2012.

The manufacturers' database sources searched and additional approaches were considered appropriate with the exception that searches in clinical trials registries such as clinicaltrial.gov were not conducted until requested. Searches in metaRegister registry of controlled trials are also recommended. Additional searches in these registers were carried out by the ERG (terms used in Appendix 1, see Section 4.1.4 below).

The searches for RCT evidence regarding the intervention (aripiprazole) and the comparators (risperidone; quetiapine; olanzapine; or in combination with lithium or valproate) were considered comprehensive. However in the searches for non-RCT evidence, the use of the heading "Epidemiologic study characteristics/" in the RCT search filter was too restrictive and not adequate to capture non-RCT evidence in this search strategy (Appendix 1b; MS page 268). This search strategy for non-RCT evidence was not clearly justified in the MS or clarification response from the manufacturer. A more appropriate method would be to apply a search filter for non-RCT evidence such as the SIGN filter.[33]

Justification for the omission of adverse events searches were not provided in the MS or explained in the clarification response. As a result, the ERG carried out supplementary searches in Medline and Embase for adverse event using previously published methods (Golder et al. 2006[34]). Further details of the supplementary searches are provided on pages 1-3 of Appendix 1. A total of 468 unique records were retrieved and reviewed by the ERG. From these records, 7 were identified as being relevant to the decision problem in the MS (see Section 4.2.3).

4.2 Clinical efficacy: systematic review of intervention

4.2.1 Inclusion/Exclusion Criteria

The inclusion and exclusion criteria used in the selection of evidence for the systematic review of clinical effectiveness were presented in the MS (pages 44/45). The MS reports that each review was performed independently by two reviewers, who then came to a consensus on the results. Details of the inclusion and exclusion criteria applied in the MS are presented in Table 2.

Table 2: Inclusion and exclusion criteria for study selection in the systematic review of clinical evidence for the treatment of acute manic and mixed episodes in bipolar I disorder in children and adolescents with aripiprazole

Inclusion Criteria
• Patients with manic or mixed episodes of bipolar I disorder • Patients aged <18 years • At least one of the interventions studied must be aripiprazole • Studies must provide sufficient detail regarding methods and results to enable the methodological quality of the study • Randomised controlled trial • Non-randomised controlled trial that still evaluates the effectiveness of interventions (acceptable study designs: prospective cohort study, retrospective chart/database review) English Language
Exclusion criteria
• Cross-Sectional or retrospective studies

No additional justification to Table 2 for the inclusion and exclusion criteria was provided in the MS. A flow diagram depicting the study selection process was provided (MS page 46).

The inclusion criteria for the review appeared reasonable and relevant to the decision problem.

4.2.2 Identified studies

Table 3: The review of clinical effectiveness evidence for aripiprazole in the MS identified the following studies

Study name and sources	Intervention	Comparator	Population
NCT00110461[21] Source: Findling *et al.*, (2009[21]) Sponsor: Otsuka	Aripiprazole Dose: 10 mg or 30 mg Duration: 4 weeks in the acute phase and for an additional 26 weeks in the extension phase	Placebo	N= 296 (10mg n=99; 30mg n=98; placebo n=99) Age: 10-17 years Diagnosis: DSM-IV bipolar I disorder, with current manic or mixed episodes, with or without psychotic features, and YMRS total score ≥ 20 at baseline.
NCT00116259[1] Source: Tramontina *et al*,. (2009[1]) Sponsor: Federal University of Rio Grande do Sul	Aripiprazole Dose: Started at 5mg, up to 20mg Duration: 6 weeks	Placebo	N= 43 (20mg n= 18; placebo n=25) Age: 8-17 years Diagnosis: DSM-IV bipolar I or II disorder comorbid with DSM-IV ADHD and YMRS score ≥ 20 at baseline).
NCT00194077[35] Source: Findling *et al.*, (2011[35]) Sponsor: University Hospitals of Cleveland	Aripiprazole Dose: Started at 0.1 mg/kg/d up to 15 mg Duration: Up to 16 weeks open label phase followed by 72 week double-blind phase	Placebo	N= 96 Age: 4-9 years Diagnosis: DSM-IV criteria for bipolar disorder (I, II, not otherwise specified, cyclothymia)

Of the 3 identified studies, only trial NCT00110461[21] was discussed in detail in the MS for the following reasons:

- Study NCT00194077[35] was excluded from the review because it did not include children over the age of 10 and therefore does not include a population for which aripiprazole is indicated.
- Study NCT00116259[1] was not discussed in detail (MS page 50) owing to it being a small study including only 23 patients receiving aripiprazole. The ERG notes that this number of patients receiving aripiprazole is discrepant to the number of patients in Table 3, which is the number cited in the Tramontina *et al.*, (2009[1]) published paper. The manufacturers stated also that, it evaluated the use of aripiprazole in a very specific population of

children and adolescents with bipolar disorder comorbid with ADHD and patients with bipolar II disorder. The ERG considers that a smaller sample size is not a valid reason for exclusion from the review and that the comorbid ADHD population included are not wholly dissimilar from the main trial NCT00110461[21] in which 51.7% are reported to have comorbid ADHD. However, the ERG considered that a mixed population of bipolar I and II disorder is a valid reason for exclusion. The manufacturers do include the trial NCT00116259[1] in the meta-analysis of the clinical evidence base for transparency and conclude that this study does not contribute substantially to the evidence base (MS page 138).

4.2.3 Studies omitted from the review

The ERG identified seven study records relating to six studies by repeating and updating the search strategy in the MS. Four of these studies were completed trials from clinicaltrials.gov which appeared relevant to the decision problem and with which Otsuka or Bristol Myers Squibb were collaborators. The ERG asked the manufacturers for clarification on whether the trials were identified and subsequently excluded. The four trials were:

i. NCT00194012[36] -"Study of Aripiprazole (Abilify) Versus Placebo in Children (5-17) With Subsyndromal Bipolar Disorder"
ii. NCT00102518[37]- "Aripiprazole Open-Label, Safety and Tolerability Study (APEX 241)"
iii. NCT00181779[38]- "Aripiprazole for the Treatment of Mania in Children and Adolescents With Bipolar Disorder"
iv. NCT00221416[39]- "An Open-Label Trial of Aripiprazole in Children and Adolescents With Bipolar Disorder"

In response to the ERG's request for clarification the manufacturers stated (pages 2-8 of clarification response) that:

i. NCT00194012 was identified in the systematic review and excluded as it investigated patients who did not meet the full criteria for bipolar disorder. No data are available for this study [The ERG notes that BMS were the collaborators for this trial and the named responsible party was Robert L. Findling].
ii. NCT00102518 was identified in the systematic review and is the extension trial to the pivotal RCT in children and adolescents with bipolar disorder (NCT00110461) and

also of the pivotal RCT in children and adolescents with schizophrenia (NCT00102063). It was originally excluded as data are only presented for the bipolar and schizophrenia patients combined. These pooled data were provided by manufacturers to the ERG in the clarification response (pages 3-6).

iii. NCT00181779 is a non-randomised, single arm study[40] that was identified in the systematic review and excluded as it included patients with bipolar II disorder and bipolar disorder not otherwise specified, as well as patients with bipolar I disorder.

The manufacturers provided a synopsis of the study to the ERG. The trial was an 8-week, open-label, prospective study of aripiprazole monotherapy in outpatients aged 6-17 years of age with a diagnosis of bipolar disorder. The aim was to assess the efficacy and tolerability of aripiprazole in this patient population. Adverse events were assessed through spontaneous self-reports, vital signs weight monitoring, and laboratory analysis.

The study enrolled 19 patients, of which 15 (79%) completed the study. Aripiprazole treatment was associated with clinically and statistically significant improvement in mean YRMS scores ($p<.0001$).

There were no statistically significant changes in weight, metabolic, or cardiovascular parameters from baseline to endpoint. The most commonly reported adverse events were sedation (57%), gastrointestinal complaints (42%), cold symptoms (32%), and headache (32%). The high rate of sedation in this sample did not account for the improvement in symptoms of mania: at study endpoint, the mean change YMRS score was not statistically different ($p=0.7$) in those who had experienced sedation ($–17.5 \pm 8.1$) than in those who had not been sedated ($–18.9 \pm 5.3$). Two cases of EPS caused the patients to discontinue medication and withdraw from the study

iv. NCT00221416[39] was a non-randomised, single arm study that was not previously identified. The manufacturers provided a synopsis for this study in their clarification response (pages 7-8). It was a 6-week, open-label, prospective study that aimed to assess the safety and efficacy of aripiprazole monotherapy in children and adolescents aged 7 to 18 years old, diagnosed with bipolar I disorder, manic or mixed episode. The study enrolled 16 patients, of which 13 (81%) completed the study. Treatment with aripiprazole was associated with significant improvement in the mean YMRS score at week 42 (mean = 6.47+/-7.8) compared with baseline (mean = 29.67+/-5.02). Aripiprazole was well tolerated, with no extrapyramidal adverse events. There was a mean weight gain of 0.99+1.4 kg ($p=0.16$). The published abstract for this study[39] reports that this increase in weight gain was statistically significant. The most common adverse events were appetite changes, nausea/vomiting and sleep problems.

The ERG considered the manufacturers' justification for the exclusion of trials NCT00194012; NCT00102518 and NCT00181779 to be satisfactory. Two further studies were identified from the updated and supplementary searches by the ERG after the clarification request was made to the manufacturer:

- Trial NCT00205699[36] is a completed study trial record for metabolic effects of antipsychotics in children and includes aripiprazole; olanzapine and risperidone (conducted by Washington University School of Medicine and National Institutes of Health). The ERG noted that this study is not specific to bipolar disorder and that there are limited outcomes which are directly relevant to this assessment such as efficacy in the treatment of mania.
- Ramos-Rios., *et al* (2009)[41] is a prospective non-randomised, single arm study investigating the effects and tolerability of aripiprazole in 12 children and adolescents with a range of psychiatric disorders. The study was not therefore specific to bipolar disorder however efficacy was measured using the CGI-S and adverse events.

In conclusion the ERG identified one single-arm study not identified by the manufacturer's search strategy (NCT00221416[39]), and two further studies (NCT00205699[36] and Ramos-Rios., *et al* (2009)[41]) which would not have qualified for inclusion into the systematic review in the MS.

4.2.4 Summary and critique of aripiprazole clinical effectiveness evidence from trials NCT00110461[21] and NCT00116259[1]

The primary data source for the trial NCT00110461[21] was a publication by Findling *et al,*. (2009[21]) and ten secondary references comprising of 9 conference abstracts (Correll *et al,*. 2008[42]; Findling *et al,*. 2007[43]; Forbes *et al,*. 2008[44]; Loze *et al,*. 2011a[45]; Loze *et al,*. 2011b[46]; NHSC 2008[47]; Pikalov *et al,*. 2009[48]; Whitehead *et al,*. 2009a[49]; Whitehead *et al,*. 2009b[50]) and one journal paper (Mankoski *et al,*. 2011[51]). The ERG learned through the clarification process that the manufacturers also sourced data on this trial from the CSR.

Trial NCT00110461[21] is a Phase III, multicentre, double-blind, randomised, placebo-controlled clinical trial recruiting 296 patients which was undertaken across 59 investigational sites in the United States between March 2005 and February 2007. The study itself was of six months duration, [redacted] and comprised a 4-week acute phase followed by a 26 week extension phase (MS page 52).

The objective of this Phase III trial was to test the safety and efficacy of two doses of aripiprazole in child and adolescent patients with bipolar I disorder, manic or mixed episode with or without psychotic features.

The inclusion and exclusion criteria for trial NCT00110461[21] (MS page 61) are shown in Table 4.

Table 4: Inclusion/exclusion criteria used in study selection for trials NCT00110461[21] and NCT00116259[1]

<table>
<tr><th colspan="2">NCT00110461</th></tr>
<tr><td colspan="2">Inclusion criteria:
Male and female subjects
Age 10 – 17 years
DSM-IV diagnosis of bipolar I disorder with current manic or mixed episodes, with or without psychotic features. Trained clinicians confirmed the primary diagnosis using the Schedule for Affective Disorders and Schizophrenia for School Aged Children: Present and Lifetime Version (K-SADS-PL).
YMRS total score ≥20 at baseline
Comorbid diagnoses were permitted including Attention Deficit Disorder (ADHD), Conduct Disorder (CD) and Oppositional Defiant Disorder (ODD) and anxiety disorders</td></tr>
<tr><td colspan="2">Exclusion criteria:
Bipolar II disorder, bipolar disorder not otherwise specified, a pervasive developmental disorder, schizophrenia, schizoaffective disorder, psychosis due to other medical conditions or concomitant medication.
Mental retardation (documented IQ<70 or clinical/social/school history suggestive of mental retardation)
DSM-IV substance or alcohol use disorder
Positive drug screen for cocaine or other substances of abuse
Sexual activity without contraceptive use
Pregnancy & lactation
Any other medical reason as determined by the investigator
Noncompliance with medication washout
Inability to swallow tablets whole
History of antipsychotic treatment resistance or neuroleptic malignant syndrome
Subjects who had made suicide attempts in the previous 6 months, had a score >3 on the Suicidal Ideation item of the Children’s Depression Rating Scale-Revised (CDRS-R), or who were determined by the investigator to be at risk of suicide
Clinically important laboratory test results, vital sign or ECG abnormalities
Diabetes mellitus
Abnormally elevated serum glucose levels
Epilepsy
History of severe head trauma
Stroke
Unstable thyroid pathology requiring treatment
Other unstable medical conditions
Prior participation in an aripiprazole study
Allergy or hypersensitivity to aripiprazole
Participation in an investigational drug trial in the past month</td></tr>
<tr><th colspan="2">NCT00116259</th></tr>
<tr><th>Inclusion criteria</th><th>Exclusion criteria</th></tr>
<tr><td>Age 8 to 17 years
DSM-IV bipolar I or II disorder comorbid with DSM-IV ADHD
Clear reports of ADHD symptom onset preceding any mood symptomatology
Acutely manic or mixed state,</td><td>Estimated IQ <70, assessed by a trained psychologist using the Wechsler Intelligence Scale for Children, Third Edition
Use of any medication 4 weeks prior to entering the study
Diagnoses of pervasive developmental disorder, schizophrenia, or substance abuse</td></tr>
</table>

defined as a YMRS score ≥20 at the baseline visit	or dependence • Severe suicide/homicide risk contraindicating outpatient treatment • Previous use of aripiprazole • Any other acute or chronic disease that might interfere in the study • Pregnancy

The manufacturers acknowledge that the inclusion and exclusion criteria for the main study NCT0011461 could have reduced the external validity of the trial population (MS page 140). The exclusion of patients who were suicidal (defined as subjects who had made suicide attempts in the previous 6 months, had a score >3 on the Suicidal Ideation item of the CDRS-R, or who were determined by the investigator to be at risk of suicide) is likely to impact the generalisability of the trial results to the clinical population. The clinical advisors to the ERG highlighted that patients who are at risk of suicide are the patients who are most relevant to this decision problem. Whilst the inclusion of suicidal patients in a placebo-controlled trial is ethically problematic, the definition of suicidal patients in the trial NCT00110461[21] could be regarded as overly precautious when compared with the studies used by the manufacturers in the indirect comparison (MS Section 5.7, page 92). The ERG has sourced the published papers of the studies included in the network meta-analysis and reviewed the included patients in these studies. In the Geller *et al.,* (2012)[52] study and the Tohen *et al*., (2007)[22] study it is reported that patients were excluded if they were at imminent or serious suicidal risk. In the Study 149 (2009)[53]; the Haas *et al.,* (2009)[54] study; and the Pavuluri *et al*., (2010)[55] study patients are not reported to be excluded on the basis of being suicidal. Although definition and description of inclusion/exclusion on the basis of suicide is lacking in these publications[53-55] it would appear that trial NCT00110461[21] used a more conservative measure of patient selection than similar trials.

Randomisation was performed according to a 1:1:1 ratio. The method of randomisation was reported in the MS to be implemented using computer generated randomisation codes according to the CSR (MS page 270) as the details were not provided in the Findling *et al.,* 2009[21] paper. Prior to randomisation, it was reported in the MS (page 51) that patients "were screened for a period of up to 28 days and if they met the entrance criteria, were randomised on Day 1 to either 10mg or 30mg of aripiprazole or to placebo". The ERG requested clarification from the manufacturers on the purpose and conduct of this 28-day screening period as the clinical advisors to the ERG commented that they were not aware that such screening periods are typically conducted in similar studies. The manufacturers responded that the screening period was 7-28 days, and was a wash-out period for prior mood-stabilising

medication, psychotropics or antidepressants. Any mood-stabilising medication, psychotropic or antidepressant had to be discontinued for at least five half-lives prior to administration of study drug. Fluoxetine in particular had to be discontinued for 28 days prior to randomisation into the study, making the screening period a maximum of 28 days (page 9 of clarification response). The manufacturers also provided the trial protocol for NCT00110461[21] at the request of the ERG. The protocol refers to this as the "washout period" only. Considering the inclusion of some patients who had previously been using antipsychotics, the ERG considers the manufacturers' clarification of the screening period as a washout period to be satisfactory. However, two of the clinical advisors to the ERG highlighted that the washout period suggests that the patients included were not likely to have severe mania which would require immediate treatment, which provides another indication that the sample may not be entirely representative of a UK clinical population.

Upon randomisation into one of the three study arms, patients reached their target dose through a forced titration schedule and proceeded with treatment at their target dose until week 4. As the number of patients who were inpatients during the study was not reported, the ERG requested clarification on this issue from the manufacturers as a forced titration schedule is presumably more difficult to enforce and monitor in an outpatient setting. The MS reported that subject evaluations took place at Day 1, Day 4 (phone call), and at Weeks 1, 2, 3 and 4 during the acute phase and [redacted] (page 52). The manufacturers responded that the proportions of patients who were inpatients during the screening phase and at baseline are not reported in the CSR (page 9 of clarification response). The data for inpatient/outpatient status for trial NCT00110461[21] were therefore not available to the ERG and as such it is possible that no patients in the NCT00110461[21] trial were inpatients.

It is reported (MS pages 52/53) that both the patient and the investigator were blinded and that [redacted]

The baseline characteristics of patients in trials NCT00110461[21] and NCT00116259[1] as reported in the MS (page 59) are presented in Table 5.

Table 5: Baseline characteristics of participants in trials NCT00110461[21] and NCT00116259[1] as presented by the manufacturers

Trial no. (acronym)	**Placebo**	**ARI 10 mg**	**ARI 30 mg**
NCT00110461 (N=296)	(n=99)	(n =98)	(n = 99)
Mean age (years) ± SD	13.3 ± 2.1	13.7 ± 2.2	13.3 ± 2.3
Gender (% male)	56.6	53.1	51.5
Mean age at onset (years) ± SD	11.9 ± 3.0	12.5 ± 3.2	12.0 ± 3.0
Mean duration of bipolar disease (years) ± SD	1.4 ± 1.9	1.3 ± 2.2	1.3 ± 2.5
YMRS total score, mean ± SD	30.7 ± 6.8	29.8 ± 6.5	29.5 ± 6.3
Weight, mean ± SD, kg	60.5 ± 17.3	63.8 ± 20.1	60.5 ± 21.5
BMI, mean ± SD, kg/m^2	23.8 ± 5.7	24.2 ± 5.4	23.7 ± 6.7
Trial no. (acronym)	**Placebo**	**ARI 20 mg**	-
NCT00116259 (N=43)	(n=25)	(n=18)	-
Mean age (years) ± SD	12.16 ± 2.75	11.72 ± 2.71	-
Gender (% male)	56	33.3	-
Mean age at bipolar disorder onset (years) ± SD	8.64 ± 3.54	7 ± 3	-
YMRS total score, mean ± SD	40.56 ± 9.01	35.94 ± 8.55	-
Weight, mean ± SD, kg	51.34 ± 18.92	48.24 ± 17.46	-

ARI= Aripiprazole; N = number of patients; SD = standard deviation; YMRS = Young Mania Rating Scale

The ERG requested clarification from the manufacturers on the number of patients in trial NCT00110461[21] who were experiencing mixed episodes, i.e., patients who were experiencing mania and depression simultaneously as this characteristic was not reported in the MS. Clinical advisors to the ERG highlighted that it needs to be stated whether the evidence applies to both mixed as well as manic episodes in bipolar I disorder. Additionally a statement received by NICE from consultation for this assessment on behalf of the Royal College of Psychiatrists stated that "it is appropriate that this assessment specifies manic and mixed episodes." Moreover, clinical advisors to the ERG stated that they are more likely to prescribe patients presenting with symptoms of irritability or agitation with an antipsychotic with a more sedative effect such as olanzapine or quetiapine (see Section 2.2). Symptoms such as irritability and agitation are more likely to reflect a mixed episode and therefore this baseline characteristic is highly relevant to the assessment of aripiprazole. The number of patients who are in mixed episodes is commonly reported in similar studies for the adult data in the EPAR (2009)[56] for aripiprazole and for the studies used in the indirect comparison in the MS.[22,52-55] The manufacturers responded that data collection relating to this clinical characteristic was not required by the study protocol and was assessed *post-hoc* (page 10 of clarification response). They state that accordingly, there is a high proportion of missing data.

As could be expected in such a population[57], a substantial proportion of patients were experiencing a mixed episode. Table 5, shows the *post hoc* data available on manic and mixed episodes provided by the manufacturers to the ERG. The number of patients who were assessed to be in a manic episode; mixed episode or unknown is relatively similar between the three study arms.

Table 6: Manic and mixed state status at baseline in study NCT00110461[21] as provided by the manufacturers at the ERG's clarification request

Current episode at baseline, n (%)	**Placebo** (n=99)	**Aripiprazole 10 mg** (n =98)	**Aripiprazole 30 mg** (n = 99)	**Total** (n=296)
Mixed	43 (43.4%)	43 (43.9%)	39 (39.4%)	125 (42.2%)
Manic	38 (38.4%)	41 (41.8%)	40 (40.4%)	119 (40.2%)
Unknown	18 (18.2%)	14 (14.3%)	20 (20.2%)	52 (17.6%)

The ERG requested clarification from the manufacturers on the number of patients in trial NCT00110461[21] who were rapid cyclers, i.e., patients who experience four or more episodes a year. The reason for this request is that the management of rapid-cycling bipolar disorder may vary slightly according to the NICE guidelines for management of bipolar disorder such as avoiding 'medication-induced switching from one pole to another, particularly with antidepressants'. The characteristic of rapid-cycling bipolar disorder is also reported in similar studies for the adult data in the EPAR for aripiprazole[26] and for the studies used in the indirect comparison in the MS (Tohen[22]; Pavuluri[55]; Geller[52]; Haas[54]). Additionally the EMA guidance[25] for clinical investigation in bipolar disorder state that it might be considered to define efficacy by a clinically relevant reduction of cycles. The manufacturers' response stated that data collection relating to this clinical characteristic was not required by the study protocol and was assessed *post-hoc* (pages 10/11 of clarification response). Accordingly, there is a relatively high proportion of missing data. Table 6, shows the *post hoc* data made available to the ERG on the number of rapid cyclers in trial NCT00110461. The number of patients who were assessed to be rapid cycling; the number who were not rapid cycling or unknown is relatively similar between the three study arms.

Table 7: Rapid cyclers in study NCT00110461[21] as provided by the manufacturers at the ERG's request for clarification

Rapid cycling*, n (%)	**Placebo** (n=99)	**Aripiprazole 10 mg** (n =98)	**Aripiprazole 30 mg** (n = 99)	**Total** (n=296)
Yes	15 (15.2%)	17 (17.4%)	13 (13.1%)	45 (15.2%)
No	51 (51.5%)	49 (50.0%)	46 (46.5%)	146 (49.3%)
Unknown	33 (33.5%)	32 (32.7%)	40 (40.4%)	105 (35.5%)

* Rapid cycling defined by DSM-IV criteria as patients who experience four or more manic, hypomanic or mixed episodes during the previous year

The ERG requested clarification from the manufacturers on whether patients in trial NCT00110461[21] were in receipt of psychotherapy and if so, whether there were any differences in the numbers receiving psychotherapy between treatment arms. The clinical advisors to the ERG stated that psychotherapy interventions such as psycho-education are evidence-based treatments; that they would be provided as adjunctive to medication in UK clinical practice and that such interventions are likely to improve adherence to medication. Additionally the EMA guidance[25] for clinical investigation in bipolar disorder state that psychotherapy, psycho-education, support or counselling may be given as supplementary treatment, but should be standardised, documented and taken into account when analysing the results. The manufacturers responded that the protocol did not preclude use of non-pharmacological therapy (page 11 of clarification response). However, this was not explicitly recorded. Information relating to the number of patients who were in receipt of psychotherapy and whether the number differed between treatment arms in trial NCT00110461[21] is therefore not available for this assessment.

4.2.5 Describe and critique the manufacturers' approach to validity assessment for each relevant trial

Trials NCT00110461[21] and NCT00116259[1] employ the Young Mania Rating Scale (YMRS) as the primary outcome measure. The MS states (page 61) that the YMRS scale is widely accepted and commonly used for measuring manic symptoms in clinical trials with children and adolescents with juvenile bipolar disorder. The clinical advisors to the ERG have stated that the use of the YMRS is appropriate for evaluating efficacy of antipsychotic medication. The ERG is satisfied that the outcome measures investigated in the MS ensure that the included studies and assessment undertaken by the manufacturers is internally valid.

Quality assessment was undertaken by the manufacturers using the suggested format in the NICE specification for manufacturer/sponsor submission of evidence template in summary form for both trials (MS page 70) and in fuller detail for trial NCT00110461[21] in Appendix 9.3 (MS pages 270/271). The table of quality assessment for trial NCT00110461[21] is presented in Table 8.

Table 8: Table of quality assessment for trial NCT00110461[21] as presented by the manufacturer

Study ID or acronym: NCT00110461[21]		
Study question	**How is the question addressed in the study?**	**Grade (yes/no/not clear/N/A)**
Was randomisation carried out appropriately?	Trial was randomised, but no details provided on how this was achieved in paper – but in CSR, says computer generated randomisation codes were used	Yes
Was the concealment of treatment allocation adequate?	Trial was double-blind, but no details provided in the paper – but the CSR says that interactive voice response system was used	Yes
Were the groups similar at the outset of the study in terms of prognostic factors, for example, severity of disease?	'Demographic and clinical characteristics...were similar for all 3 groups'. Large table of patient characteristics presented. However, "it should be noted that data on some clinical characteristics were missing for nearly a quarter of subjects"	Yes
Were the care providers, participants and outcome assessors blind to treatment allocation? If any of these people were not blinded, what might be the likely impact on the risk of bias (for each outcome)?	Trial was double-blind, but no details provided in paper – CSR says yes	Yes
Were there any unexpected imbalances in drop-outs between groups? If so, were they explained or adjusted for?	23.2% of patients discontinued in the placebo group, compared with 14.3% and 22.2% of patients in the aripiprazole 10 mg/day and 30 mg/day groups. Authors state that "study completion rates were high, and rates of discontinuation due to adverse events were low". Although a similar proportion of patients discontinued treatment in the placebo and aripiprazole 30 mg/day groups, in the PBO group the most common reason was lack of efficacy (8/23 patients) while in the aripiprazole 30 mg/day group, the second most common reason was adverse events (7/22).	No: Less dropouts in the aripiprazole 10 mg/day group – no explanation or discussion provided in text

Is there any evidence to suggest that the authors measured more outcomes than they reported?	Medication adherence and well-being	Yes
Did the analysis include an intention-to-treat analysis? If so, was this appropriate and were appropriate methods used to account for missing data?	No	No
Centre for Reviews and Dissemination (2008) Systematic reviews. CRD's guidance for undertaking reviews in health care. York: Centre for Reviews and Dissemination		

As evidenced by Table 8, the MS states that medication adherence and well-being were measured but not reported. There is therefore discrepancy in this item between Table 8 and the summary Table (MS page 70) in which the MS states that there is no evidence to suggest that the authors measured more outcomes than they reported. The ERG requested clarification on this issue and for full details of these missing outcomes from the manufacturer. The manufacturers responded that the [redacted] (page 11 of clarification response).

Clinical advisors to the ERG highlighted that ensuring medication adherence is a major challenge in the treatment of the acute phase in bipolar I disorder. This can be because of the positive aspects often typifying manic episodes, such as euphoria, subsequently leading to reluctance in patients to agree to take medication which will terminate the positive symptoms during the manic episode. Additionally patients may also be reluctant to initiate treatment and adhere to medication due to the side effect profile of antipsychotics. Under the Mental Health Act 1983[58], patients can be given medication against their will whilst under Section, which is commonly employed in the UK for manic episodes of bipolar I disorder. It is not uncommon for patients to be Sectioned under the Mental Health Act 1983 in the UK for management of acute manic episodes of bipolar I disorder. Indeed as mentioned previously in the ERG (Section 3.1) the clinical advisors stated that it would be extremely rare for patients with severe mania to be managed in the community and that such patients would be almost certainly be inpatients. As information on whether patients adhered to medication or 'compliance' data are not available, it is unclear to what extent the patients in the trial adhered to their allocated medication. This would not be the case in UK clinical practice where adherence to medication would be strictly monitored in the inpatient setting.

As evidenced in Table 6, the MS states (page 70 and page 271) that the analysis did not include an intention-to-treat (ITT) analysis. The ERG requested clarification from the manufacturers on why an ITT analysis was not undertaken. The manufacturers responded that "subjects were analysed as per randomised group, regardless of protocol violation. The efficacy ITT population was defined as all randomised subjects evaluated as per randomised group regardless of protocol violation. For the endpoints measured as "change from baseline" (including the primary endpoint) a modified ITT analysis was necessarily performed where patients must have recorded at least one post-baseline score to be included in the dataset (page 14 of clarification response). Therefore an ITT analysis of the whole randomised population in the strict sense was not applicable and was not performed." The ERG considers that whilst the MS stated that an ITT analysis was not performed, the manufacturers used a modified ITT analysis.

Comparability of trials NCT00110461[21] and NCT00116259[1]

Selection of patients

The proportion of patients screened for eligibility into the trial and the proportion of patients enrolled for inclusion varied between trials NCT00110461[21] and NCT00116259.[1] A very low proportion of patients who were screened for eligibility in trial NCT00116259[1] were included in the final study. Table 9 shows the numbers in both trials who were screened and the numbers who were subsequently included in the trials. One likely reason for the low percentage of patients enrolled in trial NCT0011629 following screening is because recruitment for the trial was performed through press releases (Tramontina *et al.*, 2009[1]), rather than through clinical referral as in trial NCT00110461.[21] Additionally, as trial NCT00110259 investigated aripiprazole for patients with bipolar I disorder and comorbid ADHD, this represents a more restrictive sample than the patients included in trial NCT00110461.[21]

Table 9: Number of patients enrolled in trials NCT00110461[21] and NCT00116259[1] after being screened for eligibility

	Number of patients screened	Number of patients enrolled	% of patients enrolled from screening
NCT00110461[21]	413	296	71.7%
NCT00116259[1]	710	43	6.1%

Attrition

The number of patients who dropped out of the trial is reported at 4 and 30 weeks (MS page 67). The percentage of patients who had dropped out of trial NCT00110461[21] at 4 weeks is

higher than the percentage of patients who had dropped out of trial NCT00116259[1] at 6 weeks. However as the overall number of subjects included in trial NCT00116259[1] is small, the numbers for attrition cannot easily be compared. The number of drop outs in studies of the comparator antipsychotics at 3 weeks (Section 4.3.4 of this report) is comparable to those seen in trial NCT00110461.[21]

Table 10: Number of drop outs at 4 weeks, 12 weeks and 30 weeks in trials NCT0011461[21] and NCT00116259[1]

NCT00110461[21] N= 296	Aripiprazole 10mg n/N (%)	Aripiprazole 30mg n/N (%)	Placebo n/N (%)
	Dropped out by 4 weeks		
	14/98 (14.3%)	22/99 (22.2%)	23/99 (23.2%)
	Dropped out by 12 weeks		
	NR*	NR*	NR*
	Dropped out by 30 weeks		
	64/98 (65.3%)	77 (77.7%)	87 (87.9%)
NCT00116259[1] N= 43	Aripiprazole 20mg n/N (%)		Placebo n/N (%)
	Dropped out by 6 weeks		
	1/18 (5.5%)		1/25 (4%)

*The EPAR for aripiprazole (2013) reports the discontinuation due to lack of efficacy or lack of tolerability at week 12 was: 10 mg: 11 (16.7%); and placebo 26 (43%). The figure for aripiprazole 30mg is not reported.

4.2.6 Outcomes

Outcome time points

The MS reports outcomes for trial NCT00110461[21] at 4 weeks (acute phase) and at 26 weeks (extension phase). Data at 12 weeks are not presented for the main analysis of YMRS but is presented in the subgroup analyses as may be relevant to the CHMP licence restriction. The MS states (page 52) that subject evaluations took place [redacted]

For trial NCT00116259[1] the time point for the primary outcome was 6 weeks.

Statistical analyses

The statistical analysis for trial NCT00110461[21] for the primary endpoint is described as an overall F-test for mean change from baseline in YMRS total score at a significance level of 0.05 (two-tailed) for the aripiprazole 10 mg, aripiprazole 30 mg and placebo groups (MS pages 64/65). The differences between groups (aripiprazole 10 mg vs. placebo and aripiprazole 30 mg vs. placebo) were investigated using a 2-tailed test at 5% significance.

Changes in scores from baseline were analysed using analysis of covariance (ANCOVA) with treatment as a factor and baseline score as a covariate at each time point. Least squares (LS)

means were used for the treatment comparisons. Two-tailed Student t-tests were used to test differences between the LS means within the ANCOVA model. The proportion of responders was analysed using chi-squared tests. The proportion of patients with clinically significant weight gain (≥ 7% increase from baseline) was tested using the Fisher exact test.

The MS states that the study was designed to have 85% power to detect a difference between aripiprazole and placebo of a -5.1 point change from baseline YMRS total score at week 4. Analyses of safety and tolerability included data from all randomised subjects who had taken at least 1 dose of study medication (safety sample). The efficacy sample included all patients in the safety sample who had at least 1 post-baseline efficacy assessment. All analyses were conducted in the LOCF dataset.

Clinical efficacy response

For trial NCT00110461[21] the primary outcome was the mean change from baseline to week 4 on the YMRS total score. This outcome was also reported as a secondary outcome at 30 weeks. For trial NCT00116259[1] the primary outcome was the mean change in YMRS from baseline to week 6. The secondary outcomes considered in the included trials NCT00110461[21] and NCT00116259[1] and reported in the MS are changes from baseline scores in the following scales:

i. Children's Global Assessment Scale (CGAS);
ii. Clinical Global Impressions Scale-Bipolar Version (CGI-BP) severity of mania, depression and overall bipolar illness;
iii. Children's Depression Rating Scale-Revised (CDRS-R) score;
iv. General Behaviour Inventory Scale (GBI) score (consisting of 20-items with 2 subscales assessing symptoms of mania/hypomania and depression, completed by both parents/guardians);
v. Attention-Deficit Hyperactivity Disorders Rating Scale (ADHD-RS-IV) score

Primary clinical efficacy outcome: Young Mania Rating Scale

The MS reports that both aripiprazole doses demonstrated statistically significant improvements over placebo in the YMRS total score at week 4, with treatment differences from placebo of -5.99 (95% CI: -8.49 to -3.50; $p<0.0001$) for the aripiprazole 10 mg arm, and -8.26 (95% CI: -10.7 to -5.77; $p<0.0001$) for the aripiprazole 30 mg arm (MS page 73). Table 9 shows the mean changes from baseline in the YMRS at 1; 2; 3; 4; and 30 weeks. Data for the main analysis of YMRS response are not presented in the MS at 12 weeks but are

presented for the subgroup analysis according to age and presence of comorbid ADHD at 12 weeks. The MS states that observed analysis was not available for this submission, but the CHMP assessment report stated that observed case analysis failed to show statistical significance for aripiprazole over placebo for both doses on all analysed efficacy endpoints at week 12. This, coupled with the high discontinuation rate, was the reason that the CHMP restricted treatment length with aripiprazole to 12 weeks (MS page 74).

Table 11: Mean changes from baseline in YMRS in the acute phase of study NCT00110461[21] as reported in the MS

		Placebo		Aripiprazole 10 mg/day			Aripiprazole 30 mg/day			p value vs. placebo
		N	Mean change from baseline	N	Mean change from baseline	Treatment difference (95% CI)	N	Mean change from baseline	Treatment difference (95% CI)	
YMRS	Week1	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	Week 2	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	Week 3	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	Week 4	92	-8.2	96	-14.2*	-5.99 (-8.49 to -3.50)	99	-16.5*	-8.26 (-10.7 to -5.77)	*p<0.0001
	Week 30	94	-8.2	96	-14.1*	[redacted]	99	-14.9*	[redacted]	*p<0.0001

The ERG requested clarification on why the number of patients in the placebo group at week 30 (MS page 321) was higher than the number of patients at week 4 (MS page 75). The manufacturers responded that "the variation in the n numbers reflects the fact that rating scales were not always completed for all subjects at all visits. For change from baseline analysis by visit, only subjects who had both baseline and post-baseline values were included in the LOCF datasets. A greater number of subjects would be expected to be included in the LOCF analysis at week 30 than at weeks 1-4 for placebo, since there was no value to impute in LOCF until a post-baseline visit was recorded" (page 15 of clarification response) The ERG is satisfied with the manufacturers' explanation regarding why more patients had a post-baseline value at week 30 than patients at week 4.

Data for the CGAS; CGI-BP; CDRS-R; GBI; ADHD-RS-IV scales are presented on page 75 of the MS for the (4 week) acute phase and page 321 of the MS for the (30 week) extension phase. The MS reports that both aripiprazole doses were also statistically significant at week 4 in the mean changes from baseline in CGAS score (p<0.0001); CGI-BP Severity scores for Mania (p<0.0001) and Overall Bipolar Illness (p<0.0001); GBI-Parent/Guardian Version

(p<0.0001) and Subject Version Mania Total score (10mg p=0.0468; 30mg p=0.0296); and the ADHD-RS-IV Total score (p<0.0001). Significant differences were not observed for the (4 week) acute phase in the CGI-BP Severity scores for depression; GBI- Patient Depression total scores or the CDRS-R score. A significant difference was observed in the 10mg arm for the GBI-Parent/Guardian version for depression (p= 0.0430) but not in the 30mg arm.

The presentation of observed case data in the MS is limited to subgroup analysis (see Tables 16 and 17 of this report).

YMRS Responders Analysis

The MS defined response rates as the percentage of patients achieving a ≥50% reduction from baseline YMRS (MS page 42) and states that response rates were significantly higher in the 10 mg and 30 mg aripiprazole arms compared with placebo at both week 4 (p<0.0001 and p=0.0074, respectively) and week 30([redacted]). Table 12 shows the number and percentage of patients who were defined as responders in the MS.

Table 12: Number of patients who were defined as responders and number of patients defined as in remission in trial NCT00110461[21] as presented in the MS

		Placebo			ARI 10 mg/day			ARI 30 mg/day			p value, ARI 10 mg vs. placebo	ARI 10 mg: 95% CI for difference (%)	p value, ARI 30 mg vs. placebo	ARI 30 mg: 95% CI for difference (%)	Relative risk (risk difference)	
		N	n	%	N	n	%	N	n	%					ARI 10 mg	ARI 30 mg
% responders (defined as ≥50% reduction from baseline YMRS total score)	Week 1	[redacted]														
	Week 2	[redacted]														
	Week 3	[redacted]														
	Week 4	92	24	26.1	96	43	44.8	99	63	63.6	0.0074	[redacted]	<0.0001	[redacted]	1.72 (18.70)	2.44 (37.5)
% in remission (defined as YMRS total score ≤12 and CGI-BP severity score for mania ≤2)	Week 4	N.R.	N.R.	5.4	N.R.	N.R.	25.0	N.R.	N.R.	47.5	p=0.0002	N.R.	<0.0001	N.R.	4.63 (19.60)	8.80 (42.1)

N = number of randomised subjects with both baseline and at least one post-baseline value; n = number of responders; NR = not reported ARI = Aripiprazole

Depression

The EMA guidance[25] for clinical investigation of bipolar disorder states that the occurrence of switching to depression should be investigated. The ERG requested clarification from the manufacturers on whether the occurrence of depression was explicitly measured. The manufacturers responded that "the effect of treatment on depressive symptoms was measured throughout the trial, though. These results were not reported in the original submission in order to focus on the effects of aripiprazole on manic symptoms, which is the indication under review. The 4-week data on depressive symptoms has been published (Findling 2009[21]), and both the 4-week and 30-week results are [provided]" (pages 11-14 of clarification response). The manufacturers provided the ERG with LOCF dataset for depression outcomes at weeks 4 and 30 using the CGI-BP severity depression score; the CDRS-R score; the GBI total score- parent guardian (depression); the GBI total scores- patient (depression) score. Table 13 shows the depression outcomes at weeks 4 and Table 14 shows the depression outcomes at week 30 for the aripiprazole 10mg, aripiprazole 30 mg and placebo arms in trial NCT00110461.[21] The ERG considers that whilst the data presented do not indicate concerns regarding the occurrence of depression for treatment with aripiprazole, the effect of aripiprazole is not explored in depth in the MS and conclusions about the effect of aripiprazole on depression are not explicitly made in the MS.

Table 13: Depression outcomes at week 4 (efficacy sample, LOCF dataset)[a,b] as presented by the manufacturers in response to the ERG's request for clarification

	Placebo (n=99)	Aripiprazole 10 mg (n =98)		Aripiprazole 30 mg (n = 99)	
	Value	**Value**	**P value vs. placebo**	**Value**	**P value vs. placebo**
CGI-BP severity score depression					
Baseline	2.8 (n=94)	2.9 (n=96)		2.9 (n=99)	
LS mean change at week 4	-0.6 (n=92)	-0.9 (n=96)		-0.9 (n=99)	
Treatment difference at week 4 (95% CI)		-0.25 (-0.54 to 0.04)	0.0878	-0.26 (-0.55 to 0.03)	0.0752
CDRS-R score					
Baseline	33.8 (n=86)	35.2 (n=91)		34.1 (n=94)	
LS mean change at week 4	-4.9 (n=85)	-7.2 (n=91)		-6.1 (n=64)	
Treatment difference at week 4 (95% CI)		-2.28 (-4.81 to 0.25)	0.0767	-1.19 (-3.69 to 1.32)	0.3515
GBI total scores - parent/guardian (depression)					
Baseline	13.4 (n=93)	13.4 (n=95)		12.4 (n=96)	
LS mean change at week 4	-3.8 (n=91)	-5.9 (n=95)		-4.1 (n=96)	
Treatment difference at week 4 (95% CI)		-2.13 (-4.20 to -0.07)	0.0430	-0.31 (-2.37 to 1.76)	0.7696
GBI total scores - patient (depression)					
Baseline	10.5 (n=93)	12.1 (n=96)		11.3 (n=96)	
LS mean change at week 4	-3.4 (n=91)	-3.4 (n=96)		-3.3 (n=96)	
Treatment difference at week 4 (95% CI)		0.07 (-1.73 to 1.86)	0.9418	0.19 (-1.61 to 1.98)	0.8377

[a]Variation in n numbers reflects rating scales not completed for all subjects
[b]A negative change signifies improvement on all scales reported here
CI, confidence interval; LOCF, last observation carried forward; LS least squares

Table 14: Depression outcomes at week 30 (efficacy sample, LOCF dataset)[a,b] as presented by the manufacturers in response to the ERG's request for clarification

	Placebo (n=99)	Aripiprazole 10 mg (n =98)		Aripiprazole 30 mg (n = 99)	
	Value	**Value**	**P value vs. placebo**	**Value**	**P value vs. placebo**
CGI-BP severity score depression					
LS mean change at week 30[26]	-0.5	-0.7	NS	-0.9	<0.05
Treatment difference at week 30 (95% CI)[27]				[redacted]	[redacted]
CDRS-R score					
LS mean change at week 30[27]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
GBI total scores - parent/guardian (depression)					
LS mean change at week 30[26]	-2.8	-5.0	<0.05	-4.1	NS
GBI total scores - patient (depression)					
LS mean change at week 30[26]	-3.2	-4.0	NS	-4.4	NS

[a]Variation in n numbers reflects rating scales not completed for all subjects
[b]A negative change signifies improvement on all scales reported here
CI, confidence interval; LOCF, last observation carried forward; LS least squares

Recurrence

Data on recurrence of manic episodes are not reported in the MS for trials NCT00110461[21] and NCT00116259[1] despite being listed in the decision problem. The MS states (page 80) that although [redacted] Page 140 of the MS stated that [redacted]. However this figure is not broken down between treatment arms. The [redacted] is reported on page 80 of the MS. It is reported that [redacted]

[redacted]

The manufacturers note (MS page 138) that the major limitation of the study design for informing the decision problem is that patients continued treatment to week 30 and therefore there are no data on relapse rates when patients stop treatment at week 12 as per the licence. The ERG requested clarification from the manufacturers on why relapse/recurrence was not measured in trial NCT00110461.[21] The manufacturers responded that "the trial was not designed to assess relapse/recurrence after discontinuation of treatment. A different design would be necessary to measure relapse/recurrence after stopping treatment, whereby all patients receive drug initially, reach a period of sustained remission, and then are randomly taken off drug and allocated to blinded placebo" (page 11 of clarification response). This justifies why the trial was not designed to measure relapse. However medication does not need to have been discontinued in order to measure recurrence as according to the EMA guidelines[25] recurrence is defined as a re-emergence of symptoms (new episode) after a time with no or minimal symptoms. As all analyses in the MS for trials NCT00110461[21] and NCT00116259[1] are performed using LOCF imputation (MS page 64), the impact of recurrence of manic and mixed episodes is unknown. The EMA guidance[25] for clinical investigation of bipolar disorder states that the risk of under- or overestimation of effect should be addressed. LOCF analysis may bias the assessment of treatment benefit because it assumes that patients who dropped out of the trial maintained treatment effect. In the assessment of manic or mixed episodes in bipolar I disorder, it is likely that patients who dropped out of the trial worsened or reverted to their baseline manic state.[59,60] However the use of LOCF data may be considered more appropriate than using the small numbers (particularly in the placebo arm) of observed cases.

Adverse Event data

The safety evidence for aripiprazole presented in the MS is limited to data from the included RCTs as searches for non-RCT evidence for adverse events by the manufacturers were not sufficient to capture non-RCT evidence (see section 4.1 of this report). The MS presents adverse event data for trial NCT00110461[21] in the acute phase (4 weeks, pages 123/124), and also for the end of the extension phase (30 weeks, pages 126/127). Changes in baseline metabolic parameters in the acute phase and extension phase are also presented. The MS reports that "in both treatment arms the majority of treatment-emergent adverse events (TEAEs) were mild or moderate in severity (page 121). There were no deaths or suicides during the study." Table 14 presents the adverse events occurring in more than 5% in any group during the acute phase of study NCT00110461[21] (MS Table B36, pages 123/124).

Table 15: Adverse events occurring in ≥5% of any group in the acute phase of the NCT00110461[21] trial (as presented by the manufacturer)

System organ/ class/adverse events	Time period 1: Acute phase (up to week 4)						
	ARI 10 mg: N (%) of patients (n = 98)	ARI 30 mg: N (%) of patients (n = 99)	Placebo: N (%) of patients (n = 97)	Relative risk for ARI 10 mg (95% CI)	Risk difference for ARI 10 mg	Relative risk for ARI 10 mg (95% CI)	Risk difference for ARI 30 mg
Mortality							
Deaths	0 (0.0)	0 (0.0)	0 (0.0)	N.A.	0.00	N.A.	0.00
Suicide	0 (0.0)	0 (0.0)	0 (0.0)	N.A.	0.00	N.A.	0.00
Adverse events							
Total AEs	72 (73.5)	75 (75.8)	57 (58.8)	1.25 (1.02 – 1.53)	14.70	1.29 (1.06 – 1.58)	17.00
Total SAEs	5 (5.1)	2 (2.0)	5 (5.2)	0.98 (0.29 – 3.27)	-0.10	0.38 (0.08 – 1.94)	-3.20
Extrapyramidal disorder	12 (12.2)	27 (27.3)	3 (3.1)	3.94 (1.15 – 13.50)	9.10	8.81 (2.77 – 28.04)	24.20
Somnolence	19 (19.4)	26 (26.3)	3 (3.1)	6.26 (1.92 – 20.44)	16.30	8.48 (2.66 – 27.08)	23.20
Fatigue	13 (13.3)	9 (9.1)	4 (4.1)	3.24 (1.09 – 20.44)	9.20	2.22 (0.71 – 6.98)	5.00
Headache	17 (17.3)	19 (19.2)	16 (16.5)	1.05 (0.56 – 1.95)	0.80	1.16 (0.64 – 2.13)	2.70
Akathisia	8 (8.2)	11 (11.1)	2 (2.1)	3.90 (0.86 – 17.71)	6.10	5.29 (1.22 – 22.96)	9.00
Nausea	9 (9.2)	12 (12.1)	4 (4.1)	2.24 (0.71 – 7.06)	5.10	2.95 (0.98 – 8.86)	8.00
Vomiting	8 (8.2)	7 (7.1)	9 (9.3)	0.88 (0.36 – 2.19)	-1.10	0.76 (0.30 – 1.97)	-2.20
Blurred vision	8 (8.2)	8 (8.1)	0 (0.0)	N.A.	8.20	N.A.	8.10
Salivary hypersecretion	3 (3.1)	8 (8.1)	0 (0.0)	N.A.	3.10	N.A.	8.10
Decreased appetite	6 (6.1)	3 (3.0)	3 (3.1)	1.97 (0.51 – 7.64)	3.00	0.97 (0.20 – 4.69)	-0.10
Dizziness	5 (5.1)	5 (5.1)	1 (1.0)	5.10 (0.59 – 44.07)	4.10	5.10 (0.59 – 43.99)	4.10
Increased appetite	2 (2.0)	5 (5.1)	3 (3.1)	0.65 (0.11 – 3.82)	-1.10	1.65 (0.41 – 6.67)	2.00
Upper abdominal pain	4 (4.1)	5 (5.1)	3 (3.1)	1.32 (0.30 – 5.74)	1.00	1.65 (0.41 – 6.67)	2.00
Dystonia	0 (0.0)	5 (5.1)	0 (0.0)	N.A.	0.00	N.A.	5.10
Exacerbation of bipolar disorder	0 (0.0)	3 (3.0)	5 (5.2)	N.A.	-5.20	0.58 (0.14 – 2.35)	-2.20
Extrapyramidal symptom categories							
Dystonic event (dystonia and muscle spasms)	0 (0.0)	7 (7.0)	2 (2.0)	N.A.	-2.00	3.50 (0.73 – 16.78)	5.00
Parkinsonism event (extrapyramidal disorder, bradykinesia and tremor)	14 (14.2)	29 (29.2)	4 (4.1)	3.46 (1.18 – 10.18)	10.10	7.12 (2.59 – 19.56)	25.10
Dyskinetic event	2 (2.0)	0 (0.0)	0 (0.0)	N.A.	2.00	N.A.	0.00

Residual event (muscle twitching)	1 (1.0)	1 (1.0)	0 (0.0)	N.A.	1.00	N.A.	1.00
Akathisia event (akathisia and psychomotor activation)	8 (8.1)	12 (12.1)	2 (2.0)	4.05 (0.86 – 19.98)	6.10	6.05 (1.36 – 26.87)	10.10
Any extrapyramidal symptom event	23 (23.5)	39 (39.4)	7 (7.2)	3.26 (1.47 – 7.26)	16.30	5.47 (2.57 – 11.64)	32.20

CI = confidence interval; N.A. = not available
Dystonic event: dystonia, emprosthotonos, muscle contractions involuntary, muscle rigidity, muscle spasms, muscle spasticity, myotonia, nuchal rigidity, oculogyration, opisthotonos, pleurothotonus, risus sardonicus, torticollis, trismus
Parkinsonism event: akinesia, asterixis, athetosis, bradykinesia, cogwheel rigidity, essential tremor, extrapyramidal disorder, freezing phenomenon, hypertonia, hypokinesia, hypokinesia neonatal, intention tremor, masked facies, Parkinson's disease, parkinsonian crisis, parkinsonian gait, parkinsonian rest tremor, parkinsonism, tremor, tremor neonatal
Akathisia event: akathisia, hyperkinesias, psychomotor hyperactivity, hyperkinesias neonatal
Dyskinetic event: ballismus, buccoglossal syndrome, choreoathetosis, clumsiness, dyskinesia, dyskinesia neonatal, dyskinesia esophageal, fumbling, on-and-off phenomena, tardive dyskinesia, head titubation
Residual event: chorea, Huntington's chorea, muscle twitching, myoclonus, clonus
Subjects with multiple adverse event terms within the same category counted only once toward the total. Subjects with extrapyramidal symptom events within multiple categories counted only once toward the total.
Adapted from European Public Assessment Reports published by the European Medicines Agency

The manufacturers noted (MS page 122) that [redacted]

The ERG notes that combining data from adverse events into a single group category, such as EPS, may not enable the impact of the adverse events within the category to be fully evaluated. This is because EPS can encompass relatively minor and transient side effects such as akathisia (an inner restlessness or inability to remain motionless) as well as tardive dyskinesia which is a serious and permanent condition characterised by involuntary movements and which is incurable.

The ERG have reviewed the FDA clinical review[61] for this study which reports the incidence of serious adverse events (SAEs) in the acute phase. This reports states that a "total of 5/98 (5.1%) subjects in the aripiprazole 10 mg arm, 2/99 (2.0%) in the aripiprazole 30 mg arm, and 5/97 (5.2%) in the placebo arm experienced SAEs during the acute phase, the majority of which were moderate or severe in intensity. The most commonly reported SAEs during the entire study were bipolar disorder (9/294 subjects; 3.1% overall) and bipolar I disorder (3/294 subjects, 1.0% overall).

Other SAEs reported during the acute phase were fatigue (1 subject in the 10 mg arm), accidental overdose (1 subject in the 10 mg arm), grand mal convulsion (1 subject in the 10 mg arm, secondary

to alcohol and cocaine overdose), aggression (2 subjects in the 10 mg arm), oppositional defiant disorder (1 subject in the aripiprazole 10 mg arm), suicidal ideation (1 subject in the 10 mg arm), and respiratory arrest (1 subject in the 10 mg arm, secondary to alcohol and cocaine overdose). The FDA clinical review[61] states that the "safety review from [trial NCT00110461[21]] did not find any unexpected serious adverse events and the patterns of common adverse events of aripiprazole remained consistent with current labelling." The ERG do not consider the frequency and nature of adverse events reported in trial NCT00110461[21] to raise concerns regarding aripiprazole treatment above other atypical antipsychotic trials.

Adverse event data for trial NCT00116259[1] are presented in graph but not numeric form (MS page 133). The manufacturers stated that aripiprazole had an acceptable safety profile but that the incidence of somnolence and EPS were increased in the aripiprazole group relative to placebo. The ERG considers these increases to be consistent of the adverse event profile of similar trials of antipsychotics.

The following conclusions regarding adverse event data are made in the MS (page 120):

- In both studies NCT00110461[21] and NCT00116259[1], aripiprazole demonstrated an acceptable safety profile in children aged 13 or older, particularly with respect to weight gain and increases in serum prolactin levels.
- Study NCT00110461[21] demonstrated that the incidence of clinically significant weight gain (≥7%) was not significantly different in the 30 mg and 10 mg aripiprazole arms compared with placebo at week 4 and remained low over time. The ERG noted however that a there was a [redacted] (MS page 122).
- There were no increases in serum prolactin level, with prolactin levels in all treatment groups falling over the duration of the 30-week study.
- In contrast, somnolence and EPS occurred more frequently in patients receiving aripiprazole than placebo.
- Study NCT00116259[1] established the same safety profile for aripiprazole as study NCT00110461[21], with no significant differences in weight gain or BMI between treatment groups, and increases in somnolence and EPS in the aripiprazole arms compared with placebo.
- The CHMP limited the indication for aripiprazole to adolescents aged 13 or over due to safety concerns in younger patients.

Subgroup analyses of trial NCT00110461

Age subgroup

The MS reports safety subgroup analysis by age at week 12 (pages 129-132). The data for mean weight change (kg) from baseline by age group (week 12) and BMI (kg/m^2), mean changes from baseline by age group (week 12) are also presented (MS pages 131/132). The MS states (page 129) that these data are the basis of the CHMP's decision to limit the indication for aripiprazole in paediatric bipolar I disorder to adolescents aged 13 and over. The number of patients in each arm of the trial who are reported to be aged between 13 and 17 , and therefore within the licensed population, is 65 in the 10mg aripiprazole arm; 59 in the 30mg aripiprazole arm and 58 in the placebo arm in both the baseline OC and LOCF analysis (Table 16). These numbers add up to 182 patients. The number of patients reported in the LOCF analysis in the 10-12 age subgroup adds up to a total of 107. Therefore as the total number of patients in the trial is 296 it is evident that not all patients are included, as seven patients are missing from this LOCF age subgroup analysis. Using the subgroup analysis total number of included patients of 289, the percentage of patients who are in the 13-17 age subgroup is 63%.

Table 16: Mean weight change (in kg) from baseline by age group (week 12) as presented in the MS

Visit/week	**10-12 years**						**13-17 years**					
	Aripip 10 mg		**Aripip 30 mg**		**Placebo**		**Aripip 10 mg**		**Aripip 30 mg**		**Placebo**	
	N	**Mean**	**N**	**Mean**	**N**	**Mean**	**N**	**Mean**	**N**	**Mean**	**N**	**Mean**
Observed Cases (OC)												
Baseline	32		40		39		66		59		58	
Week 4	26	1.2	31	1.4*	26	0.4	49	0.6	42	0.9	39	0.7
Week 12	16	2.8	16	4.0*	7	0.8	33	2.6*	25	2.1	14	0.2
LOCF												
Baseline	32		40		39		66		59		58	
Week 4	30	0.9	39	1.2	37	0.3	65	0.4	57	0.7	55	0.7
Week 12	30	2.2*	39	2.6**	37	0.4	65	1.6*	57	1.3	55	0.5
* $p < 0.05$, ** $p < 0.001$ vs. Placebo (Aripiprazole 10 mg and Aripiprazole 30 mg treatment group)												

Table 17: Body mass index (kg/m^2), mean changes from baseline by age group (week 12) as presented in the MS

Visit/week	10-12 years						13-17 years					
	Aripip 10 mg		Aripip 30 mg		Placebo		Aripip 10 mg		Aripip 30 mg		Placebo	
	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean
OC												
Baseline	32		40		39		66		59		58	
Week 4	26	0.5	30	0.5*	26	0.0	49	0.1	42	0.2	39	0.2
Week 12	16	0.9	16	1.4*	7	0.0	33	0.8*	25	0.4	14	0.0
LOCF												
Baseline	32		40		39		66		59		58	
Week 4	30	0.3	38	0.5*	37	-0.0	65	0.0	57	0.1	55	0.2
Week 12	30	0.7*	38	0.9**	37	0.0	65	0.4	57	0.3	55	0.1
* $p < 0.05$, ** $p < 0.001$ vs. Placebo (Aripiprazole 10 mg, and Aripiprazole 30 mg treatment group)												

Tables 16 and 17 show that in the 10-12 age subgroup, there were significant increases from baseline in weight and BMI measurements in the aripiprazole 30mg treatment arm at 4 and 12 weeks. There were also significant increases in weight and BMI measurements in the aripiprazole 10mg treatment arm using the LOCF analysis.

Comorbid ADHD subgroup

Table 18 presents the number of patients with and without current comorbid ADHD by age subgroup (MS page 85). Table 18 presents the LOCF data for the mean change from baseline by age group 10-12. 13-14 and 15-17 for YMRS total score (MS page 86). The MS states that the CHMP reviewed subgroup data and concluded that the presence of any comorbidity, including ADHD, did not seem to influence the YMRS changes with aripiprazole at weeks 4 and 12 (MS page 18).

Table 18: Number of patients with or without current comorbid ADHD separated by age group as presented in the MS

	10 - 12 years	13 - 14 years	15 - 17 years	Total (10 - 17 years)
Patients with current comorbid ADHD	67	33	39	139
Patients without current comorbid ADHD	22	30	40	92

It is noted by the ERG from Table 18 that the total number of patients included is 231, whilst the number of patients in the study is 296. Therefore there are 65 patients (22%) not included in the ADHD subgroup analysis presented by the manufacturers whose comorbid-ADHD status is unknown. The number reported to have comorbid ADHD in Table 17 is 139 which is 60.2% of the 231 patients

included in this analysis. However the MS states on page 139 that "NCT00110461 allowed the participation of patients with comorbidities such as ADHD (153 patients, 51.7%)". There is therefore discrepancy in the MS surrounding the number of patients in trials NCT00110461[21] with comorbid ADHD as well as a proportion of missing data on this baseline characteristic.

Table 19: Patients with and without current ADHD: mean change from baseline by age group 10-12. 13-14 and 15-17 for YMRS total score (LOCF) as presented in the MS

Visit/ Week	10 – 12 years (N=89)						13 – 14 years (N=63)						15 – 17 years (N=79)					
	ARI 10 mg		ARI 30 mg		Placebo		ARI 10 mg		ARI 30 mg		Placebo		ARI 10 mg		ARI 30 mg		Placebo	
	N	LS Mean	N	LS Mean	N	LS Mean	N	LS Mean	N	LS Mean	N	LS Mean	N	LS Mean	N	LS Mean	N	LS Mean
Current ADHD																		
Week 4	21	-15.32**	25	-15.51**	20	-2.28	7	-12.63	11	-17.57*	15	-9.43	16	-16.14	12	-15.15	11	-9.46
Week 12	21	-13.95*	25	-16.27**	21	-5.48	7	-15.61	11	-18.39*	15	-7.50	16	-17.55	12	-16.15	11	-9.13
No ADHD																		
Week 4	6	-14.96	8	-15.11*	8	-7.42	15	-12.55	7	-12.84	8	-9.74	16	-12.34	15	-14.80	9	-12.17
Week 12	6	-16.67*	8	-16.23*	8	-8.64	15	-14.98	7	-13.00	8	-8.16	16	-15.84	15	-12.32	9	-12.86
* $p < 0.05$, ** $p < 0.001$ vs. Placebo																		

The data in Table 19 indicate that the majority of statistically significant differences between aripiprazole and placebo in YRMS total score were in those with current ADHD. Trial NCT00116259[1] (Tramontina *et al*., 2009[1]) included 43 patients all of whom had comorbid ADHD. Within the study period no concomitant medication was allowed; one rationale for this study was that the "proposed mechanism of action of aripiprazole suggests that it might work for both conditions". Tramontina *et al*., 2009[1]) report that patients taking aripiprazole showed a significant reduction in YMRS scores from baseline to Week 6 compared with placebo group (27.22 vs. 19.52, F = 5.87; P =.02; effect size = 0.80; 95% CI 0.15 to 1.41). It is stated on page 85 of the MS that ADHD medication was permitted at week 12. It is possible that the YMRS response in the small number of 15 to 17-year-olds without ADHD may be mediated by nonspecific effects such as regression to the mean or ADHD treatment permitted in trial NCT00110461[21] for those with comorbid ADHD, However as the current data are *post-hoc*; from a small patient number and potentially confounded by the LOCF analyses, it means that no definitive conclusion can be made

Health-related quality of life

Health-related quality of life (HRQoL) was measured by the PQ-LES-Q in trial NCT00110461[21] although this outcome was not stated *a priori* on the clinicaltrial.gov register for the trial. The MS reports (page 82) that while the results did not reach statistical significance, both aripiprazole arms demonstrated a trend for improvement relative to placebo. HRQoL was not reported for trial NCT00116259.[1] No preference-based measures of health-related quality of life for paediatric bipolar disorder were used identified in the MS.

Role of caregiver in management of paediatric bipolar I disorder

Although not specified in the decision problem, clinical advisors to the ERG highlighted that discussion of the role of the family/ caregiver in the patient's management of their illness including medication adherence as well as identifying prodromal symptoms prior to acute episodes is lacking in the MS. Moreover, as the population included in the RCTs are more likely to be outpatients than in the UK, the impact of the family's role in management of patients' recovery is even more relevant.

Impact on family/caregiver

Although not specified in the decision problem, clinical advisors to the ERG highlighted that discussion of the impact of acute manic and mixed episodes in bipolar I disorder on the patient's family or caregiver is lacking in the MS. Page 139 of the MS acknowledges that

caregiver outcomes, such as caregiver quality of life, were not measured. The impact of the burden of this illness to the caregivers cannot be evaluated in this assessment.

4.2.7 Meta-analysis: aripiprazole versus placebo

A meta-analysis of NCT00110461[21] (pooled 10-30 mg dose) and NCT00116259[1] (20 mg dose aripiprazole) was performed (MS page 89). Results were not provided as forest plots for the meta-analysis. The manufacturers concluded from their analysis that:

- The meta-analysis found aripiprazole to still be statistically significantly superior to placebo in inducing symptomatic response (as measured by >50% change in YMRS score) at weeks 1, 2 and 4, but not at week 3.
- The meta-analysis found aripiprazole to be associated with a statistically significant higher rate of EPS than placebo, but not of somnolence.
- However, the small size of study NCT00116259[1] and the different patient population from the pivotal aripiprazole study (includes bipolar II disorders and restricted to patients with ADHD) means that the results are of limited use.

4.3 Clinical efficacy: systematic review of comparators

4.3.1 Inclusion/Exclusion Criteria

The inclusion criteria used in the selection of evidence for the systematic review of clinical effectiveness for the comparators were presented in the MS (pages 94/95). The MS reports that each review was performed independently by two reviewers, who then came to a consensus on the results. Details of the inclusion and exclusion criteria applied in the MS are presented in Table 20.

Table 20: Inclusion criteria for study selection in the systematic review of clinical evidence for the treatment of acute manic and mixed episodes in bipolar I disorder in children and adolescents with comparators

RCT Evidence Inclusion Criteria
• Patients with manic or mixed episodes of bipolar I disorder only • All patients aged ≤18 • Randomised controlled trial • At least one of the interventions studied must be an atypical antipsychotic (risperidone, quetiapine or olanzapine) other than aripiprazole • Studies must provide sufficient detail regarding methods and results to enable the methodological quality of the study to be assessed and the study's data and results must be extractable • English language only
Non RCT Evidence Inclusion Criteria
• As above except that non-randomised controlled trials included which still evaluate the effectiveness of interventions (acceptable study designs: prospective cohort study, retrospective chart/database review

No additional justification to the information provided in Table 20 for the inclusion and exclusion criteria was provided in the MS. A flow diagram depicting the study selection process was provided (MS page 96).

The inclusion criteria for the review appeared reasonable and relevant to the decision problem.

4.3.2 Identified studies

The review of clinical effectiveness evidence for comparators in the MS identified the following studies:

Table 21: Table of studies identified by the manufacturers for inclusion in the network meta-analysis of atypical antipsychotics

Study name	Intervention	Comparator	Population
Haas 2009[54]	Risperidone Dose: 0.5 mg–2.5 mg or 3-6 mg Duration: 3 weeks Design: Double blind RCT	Placebo	N= 169 (0.5 mg–2.5 mg n=50; 3-6 mg n=61; placebo n=58) Age: 10-17 years Diagnosis: DSM-IV criteria for bipolar I disorder, current episode manic or mixed (confirmed by K-SADS-PL). Patients with co-occurring ADHD or DBD were also included Patients must have had a total score of ≥20 YMRS at screening and baseline.
Study 149[53]	Quetiapine Dose: 400mg or 600mg Duration:3 weeks Design: Double blind RCT	Placebo	N= 284 (400mg n= 95; 600mg n= 98 placebo n=91) Age: 10-17 years Diagnosis: Included children and adolescents (10 to 17 years, inclusive) with mania.
Tohen 2007[22]	Olanzapine Dose: 2.5-20mg Duration: 3 weeks followed by 26 weeks extension phase Design: Double blind RCT	Placebo	N= 161 (2.5-20mg n=107; placebo n=54) Age: 13-17 years Diagnosis: All subjects met diagnostic criteria for manic or mixed bipolar episodes (with or without psychotic features) according to the DSM-IV. Subjects could be inpatients or outpatients with a total score of ≥20 on the Adolescent Structured YMRS.
Pavuluri 2010[55]	Risperidone plus placebo Dose: 0.5-2.0 mg Duration: 6 weeks Design: Double blind RCT	Divalproex plus placebo Dose: 15 mg/kg	N= 66 (risperidone n= 33; divalproex n= 33 Age: 8-18 years Diagnosis: A DSM-IV diagnosis of bipolar disorder Type I (mixed or manic)
Geller 2012[52]	Risperidone	Lithium or divalproex	N= 279 (Risperidone n=89;

	Dose: 4-6mg Duration: 8 weeks Design: Controlled, randomised, no-patient-choice parallel comparison study	sodium Doses: Lithium 1.1-1.3 mg Divalproex sodium 110-125 µg/mL	lithium n=90; divalproex sodium n=100) Age: 6.0 - 15.11 years Diagnosis: DSM-IV diagnosis of bipolar I disorder manic or mixed episode for at least 4 consecutive weeks immediately preceding baseline.

Of the 5 identified studies, the Haas 2009[54]; study 149[53] and Tohen 2007[22] studies were included in the network meta-analysis. The MS states that the Pavuluri 2010[55] study was excluded as it was a small study and it increased uncertainty in the meta-analysis (page 103). The Geller 2012[52] study was excluded because it was not placebo-controlled and increased uncertainty in the meta-analysis. The Pavuluri 2010[55] and Geller 2012[52] studies include the mood stabiliser divalproex sodium as a comparator which, as stated in the decision problem (ERG Section 3.3), is not a relevant comparator to antipsychotics. This limits the generalisability of these two studies to the five studies (Haas 2009[54]; Study 149[53]; Tohen 2007[22]; NCT00110461[21] and NCT00116259[1]) included in the network meta-analysis which compare antipsychotic treatment with placebo. The ERG notes that the trial population in the Geller 2012 study is markedly different to the other included trials in terms of the method of enrolment into the study; lower mean age; significant age difference between treatment arms; high comorbid ADHD prevalence; high number of rapid cycling bipolar disorder; discrepant definition of rapid cycling bipolar disorder; and the number of patients experiencing mixed episodes (discussed in detail in Section 4.3.4 of this report). The ERG considers the exclusion of the Geller and the NCT00116259[1] trials to be appropriate, although do not deem the reason for excluding Pavuluri to be appropriate. However, the inclusion of the Pavuluri study would have little impact on the conclusions from the network meta-analysis since it is between risperidone and divalproex sodium and would have little effect on the relative efficacies of the antipsychotics. The manufacturers conducted sensitivity analyses including NCT00116259[1]; Pavuluri 2010[55] and Geller 2012.[52]

4.3.3 Studies omitted from the review

Two studies which appeared relevant to the decision problem were identified from the updated searches performed by the ERG. One study record related to an abstract by Cubells *et al*., (2011)[62]. This was an RCT which compared risperidone and divalproex in paediatric bipolar disorder. However the abstract for the study did not specify the inclusion of bipolar I disorder patients only and may have included bipolar II patients. As this information is not

explicitly stated it is unclear whether the study would qualify for inclusion in the network meta-analysis in the MS according to their criteria.

Another study by Macmillan *et al*., (2008)[63] also compared risperidone and divalproex for paediatric bipolar disorder. However this was a retrospective chart review of children aged 5-14 and would not have qualified for inclusion in the network meta-analysis. The abstract sourced for this study record also did not specify that only bipolar I patients were included.

4.3.4 Summary and critique of effectiveness evidence from trials of comparators

The inclusion and exclusion criteria for the studies of comparator (MS page 61) are shown in Table 22.

Table 22: Inclusion/exclusion criteria used in study selection for studies used in the network meta-analysis presented in MS and modified by ERG

	Haas 2009[54]	Study 149[53]	Tohen 2007[22]	Pavuluri 2010[55]	Geller 2012[52]
Inclusion Criteria	Age 10–17 years inclusive; inpatients or outpatients; DSM-IV criteria for bipolar I disorder, current episode manic or mixed (confirmed by K-SADS-PL); medically stable as determined by investigator; co-occurring ADHD or DBD permitted; total score of ≥20 YMRS at screening and baseline	Age 10-17 years, inclusive; documented clinical diagnosis of Bipolar I mania	Aged 13-17 years; inpatients or outpatients; diagnostic criteria for manic or mixed bipolar episodes (with or without psychotic features) according to the DSM-IV; total score of ≥20 on the Adolescent Structured YMRS	Aged 8-18 years old; DSM-IV diagnosis of bipolar I disorder (mixed or manic episode); medication free or currently clinically unstable on medication justifying termination of the ineffective regimen; ADHD were included if present	Aged 6-15 years inclusive; outpatients; DSM-IV diagnosis of bipolar I disorder manic or mixed episode for at least 4 consecutive weeks immediately preceding baseline; CGAS score of 60 or less at baseline; good physical health; comorbid ADHD, ODD and conduct disorders allowed; suicidal ideation allowed is no imminent risk
Exclusion Criteria	Known intellectual impairment	Patients (female) must not be pregnant or lactating; known intolerance or lack of response to previous treatment with quetiapine; previously participated in this study	Prior nonresponse to olanzapine; treatment within the previous 30 days with an experimental medication not available for clinical use; serious suicidal risk; clinically significant abnormal laboratory values at baseline; DSM-IV-TR substance dependence (except nicotine and caffeine) within the past 30 days; treatment with a long-lasting neuroleptic within 14 days prior to randomisation	Active substance abuse based on DSM-IV criteria; serious medical problems history of allergy to risperidone or divalproex; presence of autism, non-affective psychotic disorders or any other psychiatric disorder requiring pharmacotherapy	IQ of less than 70; lifetime history of schizophrenia; pervasive developmental disorder or major medical or neurological disease; substance use dependency; alcohol or drug abuse within the past 4 weeks; pregnancy; sexually active and not using contraceptives; nursing

Baseline characteristics of included studies for comparator studies are presented in Table 23 (MS pages 101/102). As the age of patients included in the Tohen 2007 study is higher than the other studies (13-17 years), the mean age of patients is correspondingly higher than the other studies (15 years).

Table 23: Characteristics of participants in the included studies across the randomised groups as presented in the MS

Haas 2009[54] **Baseline characteristic**	**Placebo**	**Risperidone 0.5-2.5mg daily**	**Risperidone 3-6 mg daily**	**p-value**
N	N=58			
Median age (years) ± SD	13.0 (10-17)	13.0 (10-17)	13.0 (10-17)	
Gender (% male)	48	56	43	
Mean age at onset (years) ± SD				
Mean duration of bipolar disease (years) ± SD				
YMRS total score, mean ± SD	31.0 (7.5)	31.1 (6.0)	30.5 (6.9)	
Weight, mean ± SD, kg				
BMI, mean ± SD, kg/m^2				
Study 149[53] **Baseline characteristic**	**Placebo**	**Quetiapine 400 mg**	**Quetiapine 600 mg**	**p-value**
N	N=89	N=93	N=95	
Mean age (years) ± SD	13.11 (2.16)	13.15 (2.18)	13.31 (2.14)	
Gender (% male)	50.5%	57.9%	60.7%	
Mean age at onset (years) ± SD				
Mean duration of bipolar disease (years) ± SD				
YMRS total score, mean ± SD	31.3 (7.1)	30.6 (6.04)	31.7 (5.59)	
Weight, mean ± SD, kg	59.71 (18.08)	60.08 (17.83)	62.48 (19.42)	
BMI, mean ± SD, kg/m^2	23.5 (5.31)	23.38 (4.77)	24.14 (5.67)	
Tohen 2007[22] **Baseline characteristic**	**Placebo**	**Olanzapine 2.5-20.0 mg/day**		**p-value**
N	54	107		-
Mean age (years) ± SD	15.4±1.2	15.1±1.3		0.250
Gender (% male)	44.4	57.0		0.13
Mean age at onset (years) ± SD	11.5±3.1	10.9±3.3		0.331

Mean duration of bipolar disease (years) ± SD	NR	NR		NR
YMRS total score, mean ± SD	NR	NR		NR
Weight, mean ± SD, kg	NR	NR		NR
BMI, mean ± SD, kg/m^2	NR	NR		NR
Pavuluri 2010[55] **Baseline characteristic**	**Divalproex 15 mg/kg/day plus placebo**	**Risperidone 0.5-2.0 mg/day plus placebo**		**p-value**
N	33	32		-
Mean age (years) ± SD	11.23±3.50	10.47±3.18		NR
Gender (% male)	57.6%	62.5%		NR
Mean age at onset (years) ± SD	NR	NR		NR
Mean duration of bipolar disease (years) ± SD	NR	NR		NR
YMRS total score, mean ± SD	NR	NR		NR
Weight at baseline Normal, n (%) Overweight, n (%)	 29 (87.9%) 30 (93.8%)	 4 (12.1%) 2 (6.3%)		NR
BMI, mean ± SD, kg/m^2	NR	NR		NR
Geller 2012[52] **Baseline characteristic**	**Risperidone 4-6 mg/day**	**Lithium 1.1-1.3 mEq/l/day**	**Divalproex sodium 111-125 µg/ml/day**	**p-value**
N	89	90	100	-
Mean age (years) ± SD	11.0 (3.0)	9.7 (2.7)	9.7 (2.4)	NR
Gender (% male)	47.2%	58.9%	44.0%	NR
Mean age at mania episode onset (years) ± SD	5.8 (2.9)	5.0 (2.7)	5.0 (2.2)	NR
Mean duration of bipolar disease (years) ± SD	NR	NR	NR	NR
YMRS total score, mean ± SD	NR	NR	NR	NR
Weight, mean ± SD, kg	40.7 (18.4)	40.2 (17.2)	38.5 (14.9)	NR
BMI, mean ± SD, kg/m^2	19.1 (4.5)	19.6 (4.3)	19.4 (3.8)	NR

Data on the number of participants who were in mixed or manic states were not provided by the manufacturers therefore the ERG sought the relevant information from the published papers. Table 24 shows the number of patients included in the studies of comparators who were reported to be in mixed or manic states.

Table 24: Manic and mixed state status at baseline in studies for comparators

Current episode at baseline, n (%)	**Haas 2009**[54] (n=169)	**Study 149**[53] (n =277)	**Tohen**[22] (n = 161)	**Pavuluri**[55] (n= 65)	**Geller**[52] (n= 279)
Mixed	NR (64%)	5 (1.8%)	107 (66.5%)	23 (35.9%)	272 (97.5%)
Manic	NR	272 (98.2%)	NR	43 (67.2)	NR

Data on the number of participants who were rapid cyclers were not provided by the manufacturers therefore the ERG sought the relevant information from the published papers. Table 25 shows the number of patients included in the studies of comparators who were reported to be in rapid cyclers. However the definition of "rapid cyclers" is not reported consistently across studies. Only the Tohen 2007[22] study used the DSM-IV criteria of rapid cycling which is defined as patients who have experienced four or more manic, hypomanic or mixed episodes during the previous year. Geller 2012[52] defined rapid cyclers as "daily rapid cyclers" which would not fit the DSM-IV criteria. Rapid cycling is not defined in the Pavuluri 2010.[55] Study 149[53] and the Haas 2009[54] study do not report data on rapid cyclers.

Table 25: Rapid cyclers at baseline in studies for comparators

Rapid cycling, n (%)	**Haas 2009**[54] (n=169)	**Study 149**[53] (n =277)	**Tohen 2007**[22] (n = 161)	**Pavuluri 2010**[55] (n= 65)	**Geller 2012**[52] (n= 279)
Yes	NR	NR	144 (89.4%)	52 (81.2%)	277 (99.3%)†

* Rapid cycling defined by DSM-IV criteria as patients who experience four or more manic, hypomanic or mixed episodes during the previous year
†Defined as "daily rapid cycling"

4.3.5 Validity assessment for each relevant trial.

The YMRS is used as the primary efficacy outcome for the Haas 2009[54]; Study 149[53]; Tohen 2007[22]; and Pavuluri 2010[55] studies. The Geller 2012[52] study employs the Clinical Global Impressions for Bipolar Illness Improvement Mania (CGI-BP-IM) scale. The ERG is satisfied that the outcome measures investigated across the included studies for comparators in the MS ensure are internally valid.

Quality assessment was undertaken in summary form by the manufacturers using the suggested format in the NICE specification for manufacturer/sponsor submission of evidence template for the Haas 2009[54]; Study 149[53] and Tohen 2007[22] studies (MS page 103) and in full in Appendix 9.5 (MS pages 273-275). Table 26 presents a summary of quality assessment for these three included studies. Quality assessment for the Pavuluri 2010[55] and Geller 2012[52] studies is presented in summary form in appendix 9.19 (MS pages 323/324).

Table 26: Quality assessment results of the studies of comparators

Trial no. (acronym)	**Haas 2009[54]**	**Study 149[53]**	**Tohen 2007[22]**	**Pavuluri 2010[55]**	**Geller 2012[52]**
Was randomisation carried out appropriately?	Not clear	Not clear	Not clear	Not clear	Yes
Was the concealment of treatment allocation adequate?	Not clear	Not clear	Not clear	Yes	No
Were the groups similar at the outset of the study in terms of prognostic factors?	Yes	Yes	Yes	No	No
Were the care providers, participants and outcome assessors blind to treatment allocation?	Not clear	Not clear	Not clear	Yes	No
Were there any unexpected imbalances in drop-outs between groups?	Yes	Yes	Yes	Yes	Yes
Is there any evidence to suggest that the authors measured more outcomes than they reported?	No	Yes	No	No	No
Did the analysis include an intention-to-treat analysis? If so, was this appropriate and were appropriate methods used to account for missing data?	Yes.	Yes.	Yes	Yes	Yes

As evidenced by Table 26, the MS states that there is evidence that Study 149[53] measured more outcomes than they reported. All included trials also report that they did an intention to treat analysis.

Comparability of studies identified in the network meta-analysis

Selection of patients

The proportion of patients screened for eligibility into the trial and the proportion of patients enrolled for inclusion varied between studies included in the network meta-analysis. A very

low number of patients who were screened for eligibility Geller 2012[52] were included in the final study. Table 27 shows the numbers who were screened and the numbers who were subsequently included across studies in the NMA. One likely reason for the low percentage of patients enrolled in Geller 2012[52] is because recruitment for the study was performed through media advertisements as well as through clinical referral. Additionally patients included in Geller 2012[52] were all treatment naïve to antipsychotics. Study 149[53] does not report how many patients were screened for eligibility prior to enrolment. Recruitment is not described in Haas 2009[54]; Tohen 2007[22]; or Pavuluri 2010[55] however the percentage of patients enrolled in the study following screening is consistent and suggests that like trial NCT00110461, recruitment was from clinical referral.

Table 27: Number of patients screened and subsequently enrolled in studies Haas 2009[54]; Study 149[53]; Tohen 2007[22]; Pavuluri 2010[55]; Geller 2012[52]; NCT00110461[21] and NCT00116259[1]

	Number of patients screened	Number of patients enrolled	% of patients enrolled from screening
Haas 2009[54]	237	170	71.7%
Study 149[53]	NR	283	NR
Tohen 2007[22]	214	161	75.2%
Pavuluri 2010[55]	108	66	61.1%
Geller 2012[52]	5671	290	5.1%
NCT00110461[21]	413	296	71.7%
NCT00116259[1]	710	43	6.05%

Comorbid ADHD

The number of patients with comorbid ADHD in the trials included in the network meta-analysis varied between studies. Table 29 shows the number of patients reported to have comorbid ADHD in the Haas 2009[54]; Study 149[53]; Tohen 2007[22]; Pavuluri 2010[55] and Geller 2012[52] studies. Study 149[53] and Geller 2012[52] have very high numbers of comorbid ADHD in their included patients. Haas 2009[54] and Study 149[53] have similar levels of comorbid ADHD to trial NCT0011461 which was 51.7%. Pavuluri 2010[55] has a much lower percentage of patients with comorbid ADHD.

Table 28: Number of patients with comorbid ADHD in studies Haas 2009[54]; Study 149[53]; Tohen 2007[22]; Pavuluri 2010[55] Geller 2012[52]; NCT00110461[21] and NCT00116259[1]

	Patients with comorbid ADHD n/N	Patients with comorbid ADHD %
Haas 2009[54]	85/169	50.3%
Study 149[53]	124/277	44.8%
Tohen 2007[22]	159/161	98.8%
Pavuluri 2010[55]	12/65	18.5%
Geller 2012[52]	259/279	92.8%
NCT00110461[21]	NR	57.1%
NCT00116259[1]	43/43	100%

Attrition

The number of patients who dropped out of the trials identified in the network meta-analysis by the acute phase study endpoint is presented in Table 30. The levels of attrition are relatively comparable between studies Study 149[53]; Tohen 2007[22]; Pavuluri 2010[55]; Geller 2012[52] and NCT00110461[21] however the level of attrition is somewhat lower for Haas 2009[54] in the risperidone 0.5 mg–2.5 mg treatment arm. The number of patients reported to have dropped out of the NCT00116259[1] trial are low however the number of patients included in this study is small and therefore this level of attrition cannot easily be compared.

Table 29: Number of drop outs at end of the trial in studies Haas 2009[54]; Study 149[53]; Tohen 2007[22]; Pavuluri 2010[55] Geller 2012[52]; NCT00110461[21] and NCT00116259[1]

Haas 2009[54] N=169	Risperidone 0.5 mg–2.5 mg n=50	Risperidone 3-6 mg n=61	Placebo n=58
	Dropped out by 3 weeks		
	5/50 (10%)	15/61 (24.6%)	12/58 (20.7%)
Study 149[53] N=284	Quetiapine 400mg n= 95	Quetiapine 600mg n= 98	Placebo n=91
	Dropped out by 3 weeks		
	19/95 (20.0%)	18/98 (18.4%)	25/91 (27.5%)
Tohen 2007[22] N=161	Olanzapine 2.5-20mg n=107		Placebo n=54
	Dropped out by 3 weeks		
	22/107 (20.6%)		19/54 (35.2%)
Pavuluri 2010[55] N=66	Risperidone 0.5-2.0 mg plus placebo n= 33		Divalproex15 mg/kg plus placebo n= 33
	Dropped out by 6 weeks		
	6/33 (18.2%)		16/33 (48.5%)
Geller 2012[52] N=279	Risperidone 4-6mg n=89	Lithium 1.1-1.3 mg n=90	Divalproex sodium 110-125 μg/mL n=100
	Dropped out by 8 weeks		
	14/89 (15.7%)	32/90 (35.5%)	26/100 (26%)
NCT00110461[21] N= 296	Aripiprazole 10mg n/N (%)	Aripiprazole 30mg n/N (%)	Placebo n/N (%)
	Dropped out by 4 weeks		
	14/98 (14.3%)	22/99 (22.2%)	23/99 (23.2%)
	Dropped out by 12 weeks		
	NR*	NR*	NR*
	Dropped out by 30 weeks		
	64/98 (65.3%)	77 (77.7%)	87 (87.9%)
NCT00116259[1] N= 43	Aripiprazole 20mg n/N (%)	Placebo n/N (%)	
	Dropped out by 6 weeks		
	1/18 (5.5%)	1/25 (4%)	

4.3.6 Outcomes

Outcome time points

The three included trials (Haas 2009[54]; Study 149[53]; Tohen 2007[22]) report change from baseline at 3 weeks YMRS. The Pavuluri 2010[55] reported YMRS at 6 weeks and the Geller[52] study reported CGI-BP-IM at 8 weeks.

Table 30: Relevant outcomes from Haas 2009[54], Study 149[53] and Tohen 2007[22] as presented in the MS (pages 104-106)

Haas 2009[54]		Placebo		Risperidone 0.5-2.5 mg daily		Risperidone 3-6 mg daily		p value vs. placebo
		N	Value	N	Value	N	Value	
YMRS response (≥50% reduction from baseline in total YMRS)	% at Week 1	58	8.8	50	29.2	61	20.3	
	% at Week 2	58	26.3	50	57.1	61	61.7	
	% at Week 3	58	26.3	50	59.2	61	63.3	Risperidone 0.5-2.5 mg: p=0.002 Risperidone 3-6 mg: p<0.001
Discontinuation rate	n (%) at Week 3	58	12 (20.7)	50	5 (10.0)	61	15 (24.6)	
Extrapyramidal symptoms	% at Week 3	58	5	50	8	61	25	
Somnolence	n (%) at Week 3	58	11 (19)	50	21 (42)	61	34 (56)	
Clinically significant weight gain	% at Week 3	58	5.3	50	14.3	61	10	
Clinically significant increase in prolactin	% at Week 3	Males: 26 Females: 27	Males: 0 Females: 0	Males: 24 Females: 21	Males: 0 Females: 23.8	Males: 20 Females: 33	Males: 5 Females: 36.4	

Study 149[53]		Placebo		Quetiapine 400 mg		Quetiapine 600 mg		p value vs. placebo
		N	Value	N	Value	N	Value	
YMRS response	n (%) at Week 1							
	n (%) at Week 2							
	n (%) at Week 3	89	37	93	64	95	58	Quetiapine 400 mg: p=0.001 Quetiapine 600 mg: p=0.005
Discontinuation rate	n (%) at Week 3	90	25 (27.5)	95	19 (20.0)	98	18 (18.4)	
Extrapyramidal symptoms	n (%) at Week 3	90	1 (1.1)	193	7 (3.6) – quetiapine 400 mg and 600 mg pooled results			
Somnolence	n (%) at Week 3	165	14 (8.5)	340	100 (29.4)			
Clinically significant weight gain	n (%) at Week 3		0		12 - quetiapine 400 mg and 600 mg pooled results			
Clinically significant increase in prolactin	n (%) at Week 3		Males: 4 Females: 0		Males: 13.4 Females: 8.7			

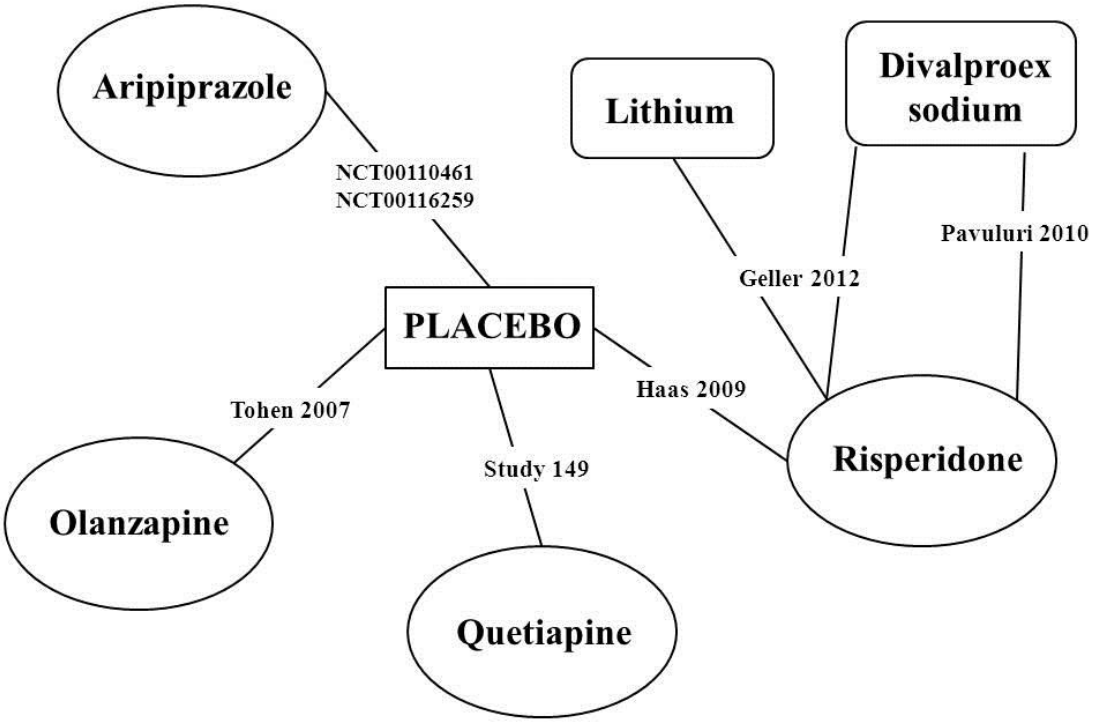
Tohen 2007[22]		Placebo		Olanzapine 2.5-20.0 mg/day				p value vs. placebo
		N	Value	N	Value			
YMRS response	n (%) at Week 1	54	11.6%	105	27.1%	-	-	<0.05
	n (%) at Week 2	54	16.8%	105	47.0%	-	-	<0.05
	n (%) at Week 3	54	22.2%	105	48.6%	-	-	0.002
Discontinuation rate	n (%)	54	35.2%	107	20.6%	-	-	NR
Any extrapyramidal symptom event	n (%) at Week 3	54	NR	107	NR	-	-	NR
Somnolence	n (%) at Week 3	54	NR	107	NR	-	-	NR
Clinically significant weight gain (≥7%)	n (%) at Week 3	54	1.9%	107	41.9%	-	-	NR
Clinically significant increase in prolactin	n (%) at Week 3	54	2.2%	107	46.7%	-	-	NR

4.4 Results of the network meta-analysis

Statistical analyses

An indirect comparison between aripiprazole; risperidone; quetiapine; and olanzapine was performed via a network meta-analysis of the trials: NCT00110461[21]; Haas 2009[54]; Study 149[53] and Tohen 2007[22] studies. Figure 1 presents the studies identified in the MS for inclusion in the indirect comparison.

Figure 1: A schematic of the network meta-analysis employed in the MS, produced by the ERG



In the manufacturers' base case analysis, the Haas 2009[54]; Study 149[53], Tohen 2007[22] and trial NCT00110461 (Findling 2009[64]) studies were included in the network meta-analysis. The MS states that the Pavuluri 2010[55] study was excluded as it was a small study and that the Geller 2012[52] study was excluded because it was not placebo-controlled and their inclusion increased uncertainty in the meta-analysis without adding further significant information (page 103). The MS additionally stated that the trial NCT00116259[1] was excluded because it evaluated the use of aripiprazole in a very specific population of children and adolescents with bipolar disorder comorbid with ADHD and patients with bipolar II disorder. The manufacturers conducted sensitivity analyses including NCT00116259[1]; Pavuluri 2010[55] and Geller 2012.[52]

The ERG considers the exclusion of the Geller 2012 and the NCT00116259[1] trials to be appropriate, although do not deem the reason for excluding Pavuluri 2010 to be appropriate. However, the inclusion of the Pavuluri 2010 study would have little impact on the conclusions from the network meta-analysis since it is between risperidone and divalproex sodium and would have little effect on the relative efficacies of the antipsychotics.

Studies which had treatment arms with different intervention doses were pooled by the manufacturers to provide an average treatment dose effect. Table 31 presents a summary of the included studies and the data used in the analysis (MS page 108). The efficacy outcomes that were analysed in the network meta-analysis were as follows:

- YMRS response (defined as ≥50% reduction in YMRS total score from baseline) at week 1
- YMRS response at week 2
- YMRS response at week 3
- Discontinuation at week 3 (all discontinuations were included, not just those for lack of efficacy)

Table 31: Summary of data used in the network meta-analysis

	YMRS week1	YMRS week2	YMRS week3	Discont. week 3	Extrapyr. symptoms	Weight gain	Prolactin increase	Somnolence
Pivotal RCTs								
Findling 2009[21]	✓	✓	✓	✓	✓	✓	✓	✓
Tohen 2007[22]	✓	✓	✓	✓		✓	✓	
Haas 2009[54]	✓	✓	✓	✓	✓	✓	✓	✓
Study 149[53]			✓	✓	✓	✓	✓	✓
Other RCTs								
Tramontina 2009[1]	✓	✓	✓	✓	✓			✓
Pavuluri 2010[55]				✓	✓			
Geller 2012[52]								✓

The statistical analysis was a Bayesian model and is described on pages 109-111 of the MS. The manufacturers justify the use of a fixed-effects model by stating that, "as a random-effects model is generally advocated if there is heterogeneity between study results and data are available in very few studies for this analysis, there is not enough evidence to support the estimation of a random effects model" (MS pages 109/110). The ERG did not consider this sufficient justification for the use of a fixed effects model and asked the manufacturers to conduct a random effects model assuming a homogenous between-study variance model which should be subject to sensitivity analyses. The ERG suggested priors for the between-study standard deviation of: U (0, 0.6); U (0, 1) and U (0, 2). The

manufacturers responded that "there is not enough evidence to support a random effects model, and therefore conducting such analysis using different priors would not provide any further useful information. The uncertainty intervals around the estimates in all cases would be very large, as the model would not have sufficient information to estimate the variability around each treatment effect." Therefore, the further analysis requested by the ERG was not conducted by the manufacturer. The ERG considers that heterogeneity in treatment effect in different studies is to be expected, unless the protocols were identical in all regards. If they were not identical, this is the rational to use random effects model instead of a fixed effects model. The purpose of the analysis is to capture plausible uncertainty about the true treatment effect and a fixed effects model would underestimate this uncertainty. The requested analyses were performed by the ERG.

The results presented were generated sampling 1,000 from a possible 50,000 CODA following a burn-in of 30,000 iterations. Following the clarification process it was ascertained that the convergence of the MCMC chains was assessed by the manufacturers through examination of the history trace plot and by assessment of the Monte Carlo standard error of the mean, which were reported to be less than 1% of the standard deviation of the posterior distribution (page 16 of clarification response). The ERG believes that calculating the Gelman-Rubin convergence statistics is a preferred approach, but that it is highly likely that convergence had occurred.

The ERG considered that the pooling of all doses of interventions as undertaken by the manufacturers may not be appropriate given the potential different side effect profile and acquisition cost of each dosage regimen. The ERG requested clarification and for the manufacturers to undertake for each of the comparator interventions a network meta-analysis assuming that different doses represent different treatment possibilities. The manufacturers responded that "unfortunately in the time permitted to respond to this letter it was not possible to programme a mixed treatment comparison to run these additional analyses. This analysis was not performed in the original submission because it is believed that patients on average would tend to receive a dose somewhere between the investigated doses, in order to balance the risk-benefit profiles of the therapies according to each individual situation" (page 18 of clarification response). The ERG considers that it is not clear what the evidence for beneficial treatment effect of the chosen dose is by pooling all doses of interventions together. By pooling different doses together, the assumption is that all doses of interventions are same in terms of safety and efficacy, an assumption which would need to be justified. The ERG also did not have the time to undertake this analysis, but caution that the results from the network meta-analysis will be affected by the randomisation ratios used within the pivotal trials, and that assuming that the YMRS response rates and discontinuation rates are similar for different doses may be inappropriate.

The ERG notes that within the network meta-analysis correlation between the trial NCT00110461[21] three-arm RCT (placebo and two doses of aripiprazole) was not appropriately handled but rather were considered as two separate two-armed trials. This would affect the relative efficacies of the two pooled doses compared with placebo. If the manufacturers had elected to have modelled the different dosing regimens of other antipsychotics then this correlation would also need to be addressed. Given that the manufacturers had pooled the results for different dosages for both risperidone and quetiapine, the ERG did not amend the lack of correlation within the NCT00110461[21] trial as the underlying problem would still remain for comparator interventions.

The manufacturers employed a continuity correction for the analyses of clinically significant weight gain and clinically significant increase in prolactin levels. This was required as a frequentist approach was undertaken, and 0.5 was added to each of the paired observation where the observed value was zero. The ERG has assessed that the change in the results were a Bayesian approach taken and a more informative prior assumed for the log odds ratio between the pair of interventions where a zero count was observed.

Clinically Efficacy Results from the Network Meta-Analysis

Only statistically significant differences from the results of the network meta-analysis from the MS are reproduced in the text of this report. Trends which are reported in the MS that do not reach statistical significance are not presented but can be found on pages 111-116 of the MS.

Response rate: YMRS Responders Analysis

The efficacy outcomes that were analysed in the network meta-analysis were as follows:

i. YMRS response (defined as ≥50% reduction in YMRS total score from baseline) at week 1
ii. YMRS response at week 2
iii. YMRS response at week 3
iv. Discontinuation at week 3 (all discontinuations were included, not just those for lack of efficacy)

The relative risks using placebo and aripiprazole as references derived from the analyses on the included studies are provided in the MS. The primary dose of interest is the pooled aripiprazole dose, but results for the 10 mg and 30 mg doses are also shown.

The YMRS response data are provided for 10-17 year olds and not the age sub group of 13-17 year-olds. The ERG requested clarification from the manufacturers on why the network meta-analysis was not performed in the 13-17 year-old sub group. The manufacturers responded that YMRS response

data were not available for this sub group, only the mean changes from baseline data at week 4 and week 12 were available (pages 15/16 of clarification response). These are the data in Table B13 of the MS (page 83) and the manufacturers state that these are not a suitable outcome for use in the mixed treatment comparison. The ERG considers this explanation to be satisfactory. For quetiapine, only 3 weeks' data is available therefore quetiapine is not included in Tables 32 and 33, but is within Table 34. The ERG notes that there is a trend for the median response compared with placebo to decline as weeks of treatment increase.

Table 32: YMRS response at week 1 as presented in the MS

	Aripiprazole pooled dose		Aripiprazole 10mg		Aripiprazole 30mg	
	median	**95% CrI**	median	95% CrI	median	95% CrI
RR versus Placebo						
Placebo	**1.00**	-	1.00	-	1.00	-
Aripiprazole	**3.47**	**1.77, 7.31**	3.50	1.69, 7.54	3.40	1.63, 7.37
Olanzapine	**2.59**	**1.20, 5.93**	2.59	1.19, 5.94	2.59	1.20, 5.94
Risperidone	**2.95**	**1.32, 7.07**	2.95	1.33, 7.08	2.94	1.32, 7.11
RR versus Aripiprazole						
Placebo	**0.29**	**0.14, 0.57**	0.29	0.13, 0.59	0.29	0.14, 0.61
Aripiprazole	**1.00**	-	1.00	-	1.00	-
Olanzapine	**0.74**	**0.27, 2.01**	0.74	0.27, 2.08	0.76	0.27, 2.16
Risperidone	**0.85**	**0.31, 2.34**	0.84	0.30, 2.43	0.87	0.31, 2.51

Table 33: YMRS response at week 2 as presented in the MS

	Aripiprazole pooled dose		Aripiprazole 10 mg		Aripiprazole 30 mg	
	median	**95% CrI**	median	95% CrI	median	95% CrI
RR versus Placebo						
Placebo	**1.00**	-	1.00	-	1.00	-
Aripiprazole	**2.63**	**1.83, 3.83**	2.26	1.45, 3.43	2.98	2.07, 4.30
Olanzapine	**2.68**	**1.66, 4.32**	2.68	1.66, 4.30	2.68	1.65, 4.30
Risperidone	**2.62**	**1.72, 3.93**	2.61	1.71, 3.94	2.61	1.71, 3.93
RR versus Aripiprazole						
Placebo	**0.38**	**0.26, 0.55**	0.44	0.29, 0.69	0.34	0.23, 0.48
Aripiprazole	**1.00**	-	1.00	-	1.00	-
Olanzapine	**1.02**	**0.61, 1.67**	1.19	0.68, 2.09	0.90	0.55, 1.46
Risperidone	**0.99**	**0.62, 1.57**	1.16	0.69, 1.96	0.87	0.55, 1.37

Table 34: YMRS response at week 3 as presented in the MS

	Aripiprazole pooled dose		Aripiprazole 10 mg		Aripiprazole 30 mg	
	median	**95% CrI**	median	95% CrI	median	95% CrI
RR versus Placebo						
Placebo	**1.00**	-	1.00	-	1.00	-
Aripiprazole	**2.39**	**1.76, 3.24**	2.13	1.45, 2.99	2.63	1.94, 3.53
Olanzapine	**2.12**	**1.39, 3.11**	2.12	1.39, 3.12	2.12	1.39, 3.12
Risperidone	**2.44**	**1.70, 3.37**	2.43	1.71, 3.37	2.44	1.71, 3.37
Quetiapine	**1.90**	**1.38, 2.5**	1.90	1.38, 2.55	1.90	1.39, 2.54
RR versus Aripiprazole						
Placebo	**0.42**	**0.31, 0.57**	0.47	0.33, 0.69	0.38	0.28, 0.52
Aripiprazole	**1.00**	-	1.00	-	1.00	-
Olanzapine	**0.89**	**0.56, 1.35**	1.00	0.61, 1.62	0.81	0.52, 1.22
Risperidone	**1.02**	**0.69, 1.47**	1.15	0.75, 1.77	0.93	0.63, 1.33
Quetiapine	**0.79**	**0.54, 1.15**	0.89	0.59, 1.38	0.72	0.50, 1.04

These data indicate that aripiprazole is statistically significantly more likely to produce a YMRS response than placebo, although there are no significant differences when comparing aripiprazole with olanzapine, quetiapine or risperidone.

Table 35: Discontinuation at week 3 as presented in the MS

	Aripiprazole pooled dose		Aripiprazole 10 mg		Aripiprazole 30 mg	
	median	**95% CrI**	median	95% CrI	median	95% CrI
RR versus Placebo						
Placebo	**1.00**	-	1.00	-	1.00	-
Aripiprazole	**0.73**	**0.41, 1.24**	0.55	0.24, 1.10	0.89	0.47, 1.55
Olanzapine	**0.54**	**0.28, 0.99**	0.54	0.28, 0.99	0.54	0.28, 0.99
Risperidone	**0.88**	**0.46, 1.61**	0.88	0.46, 1.62	0.88	0.46, 1.60
Quetiapine	**0.70**	**0.43, 1.11**	0.70	0.43, 1.11	0.70	0.43, 1.11
RR versus Aripiprazole						
Placebo	**1.37**	**0.80, 2.43**	1.82	0.91, 4.12	1.13	0.64, 2.13
Aripiprazole	**1.00**	-	1.00	-	1.00	-
Olanzapine	**0.75**	**0.32, 1.72**	1.00	0.38, 2.76	0.61	0.26, 1.49
Risperidone	**1.21**	**0.52, 2.78**	1.61	0.61, 4.48	0.99	0.42, 2.39
Quetiapine	**0.96**	**0.46, 2.01**	1.28	0.54, 3.27	0.79	0.37, 1.74

These data indicate that there were no significant differences when comparing aripiprazole with olanzapine, quetiapine, risperidone or placebo.

Safety Results from the Network Meta-Analysis: Adverse events

The adverse events that were analysed in the network meta-analysis were as follows:

i. Extrapyramidal symptoms
ii. Clinically significant increase in weight gain
iii. Clinically significant increase in prolactin
iv. Somnolence

Table 36: Extrapyramidal symptoms: Risk ratios versus placebo and aripiprazole as presented in the MS

	Aripiprazole pooled dose		Aripiprazole 10 mg		Aripiprazole 30 mg	
	median	**95% CrI**	median	95% CrI	median	95% CrI
RR versus Placebo						
Placebo	**1.00**	-	1.00	-	1.00	-
Aripiprazole	**5.28**	**2.65, 11.74**	3.71	1.66, 8.87	6.99	3.43, 15.51
Risperidone	**3.73**	**1.28, 13.84**	3.73	1.27, 13.84	3.72	1.28, 13.78
Quetiapine	**3.79**	**0.67, 44.77**	3.80	0.66, 44.12	3.80	0.67, 43.82
RR versus Aripiprazole						
Placebo	**0.19**	**0.09, 0.38**	0.27	0.11, 0.60	0.14	0.06, 0.29
Aripiprazole	**1.00**	-	1.00	-	1.00	-
Risperidone	**0.71**	**0.20, 2.93**	1.01	0.26, 4.47	0.53	0.15, 2.18
Quetiapine	**0.72**	**0.12, 8.38**	1.04	0.15, 12.47	0.54	0.09, 6.00

These data indicate that aripiprazole is significantly more likely to cause extrapyramidal symptoms than patients on placebo. There were no significant differences when comparing aripiprazole with olanzapine, quetiapine, risperidone or placebo. Data on somnolence were not reported for olanzapine; the manufacturers assumed that this drug would have the same rate of somnolence as the risperidone which had the lowest rate of those antipsychotics where data existed.

Table 37: Clinically significant weight gain: Risk ratios versus placebo and aripiprazole as presented in the MS

	Aripiprazole pooled dose		Aripiprazole 10 mg		Aripiprazole 30 mg	
	median	**95% CrI**	median	95% CrI	median	95% CrI
RR versus Placebo						
Placebo	**1.00**	**-**	1.00	-	1.00	-
Aripiprazole	**2.13**	**0.76, 7.20**	0.98	0.22, 4.24	3.22	1.1, 10.91
Olanzapine	**26.44**	**7.46, 130.3**	26.30	7.49, 127.6	26.35	7.49, 128.4
Risperidone	**2.54**	**0.78, 10.78**	2.53	0.78, 10.74	2.55	0.78, 10.62
Quetiapine	**23.54**	**3.92, 217.9**	22.67	3.8, 195.6	22.32	3.79, 196.8
RR versus Aripiprazole						
Placebo	**0.47**	**0.14, 1.32**	1.02	0.24, 4.55	0.31	0.09, 0.91
Aripiprazole	**1.00**	**-**	1.00	-	1.00	-
Olanzapine	**12.52**	**2.31, 76.22**	27.57	4.12, 216.2*	8.24	1.53, 50.23
Risperidone	**1.19**	**0.22, 6.94**	2.65	0.39, 20.21	0.79	0.15, 4.61
Quetiapine	**11.1**	**1.30, 116.1**	23.99	2.34, 297.00*	6.98	0.84, 72.47

These data indicate that there were no significant differences in the risk of experiencing a clinically significant increase in weight between aripiprazole and placebo. However, aripiprazole (pooled dose) was significantly less likely to induce clinically significant weight gain than olanzapine and quetiapine.

Table 38: Clinically significant increase in prolactin: Risk ratios versus placebo and aripiprazole as presented in the MS

	Aripiprazole pooled dose		Aripiprazole 10 mg		Aripiprazole 30 mg	
	median	**95% CrI**	median	95% CrI	median	95% CrI
RR versus Placebo						
Placebo	**1.00**	**-**	1.00	-	1.00	-
Aripiprazole	**0.20**	**0.01, 2.43**	0.11	0, 1.93	0.40	0.01, 4.81
Olanzapine	**33.83**	**9.15, 186.1**	32.82	9.06, 169.8	33.47	9.09, 185.2
Risperidone	**25.88**	**3.82, 291.8**	24.84	3.78, 261	25.81	3.8, 308.1
Quetiapine	**5.97**	**1.72, 32.14**	5.97	1.72, 31.87	5.97	1.72, 32.36
RR versus Aripiprazole						
Placebo	**4.89**	**0.41, 140.3**	8.74	0.52, 2179	2.47	0.21, 72.42
Aripiprazole	**1.00**	**-**	1.00	-	1.00	-
Olanzapine	**175.70**	**10.86, 6414**	310.70	12.84, 85630	88.75	5.48, 3267
Risperidone	**139.80**	**5.52, 7202**	251.20	6.82, 82530	71.13	2.84, 3736
Quetiapine	**31.22**	**1.81, 1191**	57.86	2.33, 16060	15.85	0.93, 618.5

These data indicate that there were no significant differences in the risk of a clinically significant increase in prolactin between aripiprazole and placebo. Patients on aripiprazole are, however, significantly less likely to experience this than patients on olanzapine, risperidone and quetiapine

Table 39: Somnolence: Risk ratios versus placebo and aripiprazole as presented in the MS

	Aripiprazole pooled dose		Aripiprazole 10 mg		Aripiprazole 30 mg	
	median	**95% CrI**	median	95% CrI	median	95% CrI
RR versus Placebo						
Placebo	**1.00**	**-**	1.00	-	1.00	-
Aripiprazole	**5.85**	**2.75, 14.10**	5.21	2.21, 13.27	6.41	3.02, 14.87
Risperidone	**3.39**	**1.87, 6.27**	3.40	1.88, 6.25	3.40	1.88, 6.24
Quetiapine	**3.47**	**1.90, 6.69**	3.49	1.91, 6.76	3.48	1.89, 6.73
RR versus Aripiprazole						
Placebo	**0.17**	**0.07, 0.36**	0.19	0.08, 0.45	0.16	0.07, 0.33
Aripiprazole	**1.00**	**-**	1.00	-	1.00	-
Risperidone	**0.58**	**0.23, 1.38**	0.65	0.24, 1.68	0.53	0.21, 1.24
Quetiapine	**0.59**	**0.23, 1.43**	0.67	0.25, 1.76	0.54	0.22, 1.29

These data indicate that aripiprazole was significantly more to cause somnolence than placebo. There were no significant differences compared with risperidone and quetiapine. Data on somnolence were

not reported for olanzapine; the manufacturers assumed that this drug would have the same rate of somnolence as the risperidone which had the lowest rate of those antipsychotics where data existed.

Comparison between the results for aripiprazole obtained through the network meta-analyses and through standard meta-analyses

Table 40 details the results from both direct meta-analyses and network meta-analyses. It is seen that the two sets of data are similar. This result is not unexpected as there were no closed loops within the network meta-analysis and thus the results from the direct meta-analyses would be changed marginally by the additional trials. The same logic applies to the data presented in Table 41.

Table 40: Comparison of the direct and network meta-analysis results for aripiprazole versus placebo: response rates

Aripiprazole pooled dose versus Placebo	Direct meta-analysis RR (95% CI)	Network meta-analysis RR (95% CrI)
YMRS response at week1	3.54 (1.56, 8.00)	3.47 (1.77, 7.31)
YMRS response at week 2	2.74 (1.71, 4.39)	2.63 (1.83, 3.83)
YMRS response at week 3	2.77 (1.73, 4.44)	2.39 (1.76, 3.24)
Discontinuation at week 3	0.71 (0.40, 1.26)	0.73 (0.41, 1.24)

Adverse event data

The MS also presents the results of the network meta-analysis for safety outcomes but the data are limited to the adverse events used in the economic model which are: EPS; clinically significant weight gain; clinically significant increase in prolactin; and somnolence (pages 114-116). This analysis presents the probabilities of events occurring in these four outcomes.

Table 41: Comparison of the direct and network meta-analysis results for aripiprazole versus placebo: adverse events

Aripiprazole pooled dose versus Placebo	Direct meta-analysis RR (95% CI)	Network meta-analysis RR (95% CrI)
Extrapyramidal symptoms	4.36 (2.08, 9.17)	5.28 (2.65, 11.74)
Clinically significant weight gain	1.97 (0.68, 5.73)	2.13 (0.76, 7.20)
Clinically significant increase in prolactin	0.25 (0.02, 2.68)	0.20 (0.01, 2.43)
Somnolence	7.39 (2.36, 23.17)	5.85 (2.75, 14.10)

The ERG requested the manufacturers to provide data from all studies used in the mixed treatment comparison for all adverse events. In response to the ERG's request the manufacturers provided tables of adverse event data for the five studies of the comparator drugs as reported in the published studies Haas 2009; Study 149; Tohen 2007; Pavuluri 2010; and Geller 2012 (pages 18-23 of clarification response). A preliminary investigation of the data suggests that the most frequently occurring adverse events appear relatively consistent across the studies.

Sensitivity analyses

Sensitivity analyses were conducted by the manufacturers in which the Pavuluri 2010[55] Geller 2012[52] and the trial NCT00116259[1] studies, were included in the network meta-analysis. For YMRS response and discontinuation, the inclusion of the trials was less favourable to aripiprazole in relation to the median value. There was no marked impact on EPS and the results became more favourable to aripiprazole for somnolence. These results are shown on page 303-311 in the MS. We have reproduced the results for YMRS response at week 3 and compared with the base case analysis in Table 42.

Table 42: YMRS response at week 3 pooled with sensitivity analysis

	Base case analysis		Sensitivity analyses	
	median	**95% CrI**	median	95% CrI
RR versus aripiprazole				
Placebo	**0.42**	**0.31, 0.57**	0.51	0.39, 0.67
Aripiprazole	**1.00**	-	1.00	-
Olanzapine	**0.89**	**0.56, 1.35**	1.01	0.67, 1.46
Risperidone	**1.02**	**0.69, 1.47**	1.14	0.8, 1.57
Quetiapine	**0.79**	**0.54, 1.15**	0.92	0.65, 1.27

4.5 Additional clinical work conducted by the ERG

The ERG repeated the network meta-analysis using a random effects model. Three priors for the between-study standard deviation of: U (0, 0.6); U (0, 1) and U (0, 2) were assessed. The latter two were discarded as these indicated that some of the antipsychotics may be no more efficacious than placebo, a conclusion that was not supported by our clinical advisors. This resulted in only the U(0, 0..6) prior being considered. Where observed values were zero (clinically significant weight gain and clinically significant prolactin levels) a normal distribution of mean 0 and variance of 10 was assumed plausible having additionally evaluated a normal distribution of mean 0 and variance of 100.

For reasons of expedition and clarity only six end points were considered: YMRS response scores at Week 3; discontinuation at Week 3; rates of EPS; rates of somnolence; rates of clinically significant weight gain; and rates of clinically significant increase in prolactin.

The results are presented in Tables 43 to 48. As expected, the point estimates are similar between the random effects models compared with the fixed effects models although in all cases the uncertainty was wider in the random effects model.

Table 43: Random effects versus fixed effects model: Risk ratios versus placebo YMRS response at week 3 with prior for between study standard deviation τ~ U(0,0.6) using an odds ratio model

	ERG random effect model		Manufacturers' fixed effect model	
	median	95% CrI	median	95% CrI
Placebo	1.00	-	1.00	-
Aripiprazole (pooled dose)	2.37	1.44, 3.40	2.39	1.76, 3.24
Olanzapine	2.11	1.20, 3.24	2.12	1.39, 3.11
Risperidone	2.46	1.38, 3.57	2.44	1.70, 3.37
Quetiapine	1.88	1.10, 2.91	1.90	1.38, 2.5

Median of the between study standard deviation is 0.30 with 95% CrI (0.021, 0.58). The model fitted the data well, with the residual deviance close to the total number of data points included in the analysis (7.89 vs 8).

Table 44: Random effects versus fixed effects model: Risk ratios versus placebo for discontinuation at week 3 with prior for between study standard deviation τ~ U(0,0.6) using an odds ratio model

	ERG random effect model		Manufacturers' fixed effect model	
	median	95% CrI	median	95% CrI
Placebo	1.00	-	1.00	-
Aripiprazole (pooled dose)	0.72	0.31, 1.49	0.73	0.41, 1.24
Olanzapine	0.53	0.20, 1.26	0.54	0.28, 0.99
Risperidone	0.89	0.36, 1.94	0.88	0.46, 1.61
Quetiapine	0.69	0.32, 1.38	0.70	0.43, 1.11

Median of the between study standard deviation is 0.30 with 95% CrI (0.012, 0.59). The model fitted the data well, with the residual deviance close to the total number of data points included in the analysis (8.14 vs 8).

Table 45: Random effects versus fixed effects model: Risk ratios versus placebo for extrapyramidal symptoms with prior for between study standard deviation τ~ U(0,0.6) using an odds ratio model

	ERG random effect model		Manufacturers' fixed effect model	
	median	95% CrI	median	95% CrI
Placebo	1.00	-	1.00	-
Aripiprazole	5.22	2.16, 14.25	5.28	2.65, 11.74
Risperidone	3.80	1.12, 17.83	3.73	1.28, 13.84
Quetiapine	4.39	0.64, 48.02	3.79	0.67, 44.77

Median of the between study standard deviation is 0.29 with 95% CrI (0.012, 0.58). The model fitted the data well, with the residual deviance close to the total number of data points included in the analysis (6.13 vs 6).

Table 46: Random effects versus fixed effects model: Risk ratios versus placebo for clinically significant weight gain with prior for between study standard deviation $\tau \sim U(0, 0.6)$ and normal prior N(0,10) for the log odds ratios for Olanzapine and Quetiapine using an odds ratio model with no continuity correction

	ERG random effect model		Manufacturers' fixed effect model	
	median	95% CrI	median	95% CrI
Placebo	1.00	-	1.00	-
Aripiprazole	2.17	0.65, 9.16	2.13	0.76, 7.20
Olanzapine	22.05	5.97, 90.5	26.44	7.46, 130.3
Risperidone	2.46	0.64, 13.59	2.54	0.78, 10.78
Quetiapine	21.53	4.10, 114.2	23.54	3.92, 217.9

Median of the between study standard deviation is 0.28 with 95% CrI (0.020, 0.58). The model fitted the data well, with the residual deviance close to the total number of data points included in the analysis (8.06 vs 8).

Table 47: Random effects versus fixed effects model: Risk ratios versus placebo for clinically significant increase in prolactin with prior for between study standard deviation τ~ U(0,0.6) and normal prior N(0,10) for the log odds ratios for Olanzapine and Risperidone using an odds ratio model with no continuity correction

	ERG random effect model		Manufacturers' fixed effect model	
	median	95% CrI	median	95% CrI
Placebo	1.00	-	1.00	-
Aripiprazole	0.21	0.0066, 2.74	0.20	0.01, 2.43
Olanzapine	26.58	7.27, 128.2	33.83	9.15, 186.1
Risperidone	29.24	4.39, 219.1	25.88	3.82, 291.8
Quetiapine	5.95	1.44, 32.50	5.97	1.72, 32.14

Median of the between study standard deviation is 0.29 with 95% CrI (0.013, 0.58). The model fitted the data well, with the residual deviance close to the total number of data points included in the analysis (7.96 vs 8).

Table 48: Random effects versus fixed effects model: Risk ratios versus placebo for somnolence with prior for between study standard deviation $\tau \sim U(0, 0.6)$ using an odds ratio model

	ERG random effect model		Manufacturers' fixed effect model	
	median	95% CrI	median	95% CrI
Placebo	1.00	-	1.00	-
Aripiprazole	5.82	2.28, 14.37	5.85	2.75, 14.10
Risperidone	3.36	1.42, 7.14	3.39	1.87, 6.27
Quetiapine	3.53	1.47, 7.45	3.47	1.90, 6.69

Median of the between study standard deviation is 0.29 with 95% CrI (0.013, 0.58). The model fitted the data well, with the residual deviance close to the total number of data points included in the analysis (6.19 vs 6).

Comparison of relative risks and odds ratios

Despite the manufacturers performing an odds ratio model the results were presented in terms of relative risks, which were converted from the odds ratios, as relative risks were used within the manufacturers' economic model. This methodology, whilst not as efficient as using the odds ratios directly, is not thought to bias the results calculated by the manufacturers. However, the use of the relative risks in an alternative setting where there is a different baseline would result in an error being made. For this reason the ERG present the odds ratios calculated from a random effects model, with a prior for the between study standard deviation of U(0,0.6).

Table 49: The odds ratios versus placebo associated with the ERG random effects model with prior for between study standard deviation τ~ U(0,0.6)

	Median (95% CrI)					
OR vs placebo	YMRS response (Wk 3)	Discontinuation (Wk 3)	Extrapyramidal symptoms	Clinically significant weight gain	Clinically significant increase in prolactin levels	Somnolence
Aripiprazole (pooled dose)	4.36 (1.71, 10.95)	0.66 (0.25, 1.74)	6.02 (2.25, 20.53)	2.21 (0.65, 9.95)	0.21 (0.01, 2.79)	10.07 (2.59, 48.19)
Olanzapine	3.29 (1.31, 9.44)	0.46 (0.16, 1.37)	4.21 (1.12, 24.25)	36.91 (7.23, 436.32)	40.33 (8.59, 521.69)	4.31 (1.46, 11.61)
Risperidone	4.70 (1.62,14.10)	0.86 (0.28, 2.60)	4.98 (0.63, 110.73)	2.52 (0.64, 15.50)	45.81 (4.67, 1701.13)	4.50 (1.56, 13.40)
Quetiapine	2.64 (1.14, 6.96)	0.63 (0.26, 1.54)	Assumed equal to risperidone	36.97 (4.63, 928.90)	6.46 (1.45, 48.19)	Assumed equal to risperidone

4.6 Network meta-analysis conclusions

The key clinical evidence in this submission comes from a network meta-analysis of the pivotal RCTs for each of the four atypical antipsychotics (MS page 89). The only adverse events reported in the meta-analysis are EPS, weight gain, prolactin increase and somnolence.

The manufacturers summarised the network meta-analysis as follows:

- "There were no statistically significant differences in YMRS response rates at weeks 1-3 between the atypical antipsychotics, although there was a trend for aripiprazole to have greater efficacy compared with all others at week 1 and quetiapine and olanzapine at week 3.
- Aripiprazole-treated patients were significantly less likely to experience clinically significant weight gain than patients on olanzapine (RR olanzapine vs. aripiprazole: 12.52 [95%CrI 2.31-76.22] and quetiapine (RR quetiapine vs. aripiprazole: 11.1 [95% CrI 1.30-116.1]) at study endpoint.
- Aripiprazole-treated patients were significantly less likely to experience a clinically significant increase in prolactin than patients on olanzapine (RR olanzapine vs. aripiprazole: 175.70 [95% CrI 10.86-6414]), risperidone (RR risperidone vs. aripiprazole: 139.80 [95% CrI 5.52-7202]) or quetiapine (RR quetiapine vs. aripiprazole: 31.22 [95% CrI 1.81-1191]).
- There were no significant differences between aripiprazole and the other atypical antipsychotics where data were available in terms of EPS (RR vs. aripiprazole for risperidone: 0.71 [95% CrI 0.20-2.93]; for quetiapine: 0.72 [95% CrI 0.12-8.38]) and somnolence rates (RR vs. aripiprazole for risperidone: 0.58 [95% CrI 0.23-1.38]; for quetiapine: 0.59 [95% CrI 0.23-1.43])."

The ERG considers that it is not clear what the evidence for beneficial treatment effect of the chosen dose is by pooling all doses of interventions together. By pooling different doses together, the assumption is that all doses of interventions are the same in terms of both safety and efficacy. This assumption requires justification. Overall the ERG recommends that a mixed treatment comparison should treat different doses as different treatment possibilities.

The additional analyses undertaken by the ERG, which included using a random effects model and a Bayesian adjustment for when zero values were observed, did not alter these conclusions. However it was noted that the uncertainty was wider in the ERG's analyses than in the manufacturers' analyses.

5 ECONOMIC EVALUATION

5.1 ERG comment on manufacturers' review of cost-effectiveness evidence

The manufacturers did not identify any relevant economic evaluations. The manufacturers' search strategy was based on a previously published systematic review in order to identify all relevant cost-effectiveness information available for the treatment of acute manic and mixed episodes in bipolar I disorder in children and adolescents.[28]

The MS reports that 6694 records were found. As with the systematic review for clinical efficacy (Section *4.1*) searches were limited to January 2012. The ERG was able to repeat and update the database searches until January 2013 A total of 7056 records were retrieved, of which, 955 were in 2012 which represents a significant number of records that were missed by the manufacturers searches that were conducted up to January 2012. Whilst the quality of life terms were comprehensive, the cost filter was somewhat restrictive, and the ERG recommends the use of a sensitive filter such as SIGN.[33] The ERG does not believe that any additional relevant studies were missed by the manufacturer's cost-effectiveness review.

5.2 Summary and critique of manufacturers submitted economic evaluation by the ERG

5.2.1 NICE Reference Case checklist

Table 50: Consistency of the manufacturers' economic evaluation with the NICE Reference Case[65]

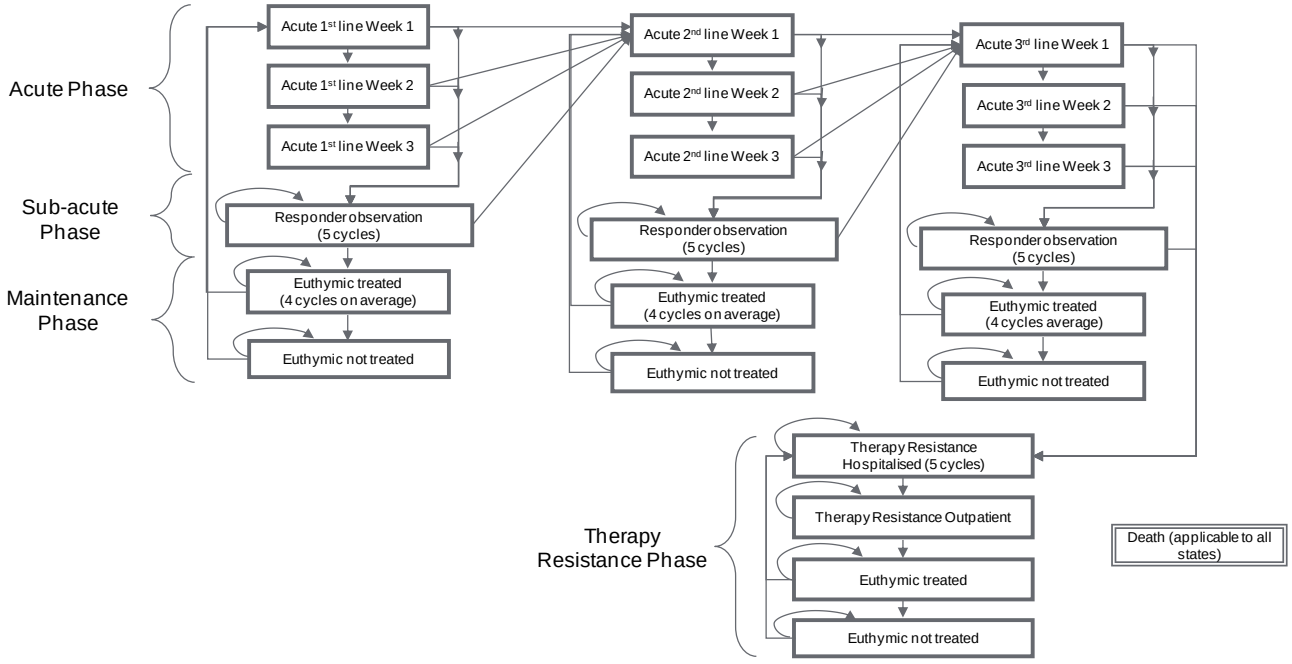
Factor	Consistent with the NICE reference case?	ERG comment
Decision problem	Yes	
Comparator	Partly	The NICE scope lists lithium and valproate; either on their own or in combination with antipsychotics. These are not included as comparators in the MS. However, the ERG's clinical advisors stated that these are rarely used (either on their own or in combination).
Perspective on costs	Yes	The perspective of the NHS and PSS was adopted.
Perspective on outcomes	Partly	Neither depression nor the potential adverse effects of prolactin increase were modelled. All other relevant health effects were included.
Type of economic evaluation	Yes	Cost-effectiveness analysis undertaken and expressed in terms of the incremental cost per QALY gained.
Synthesis of evidence on outcomes	Partly	Systematic review conducted, but only until January 2012.
Measure of health effects	Yes	Health effects measured as QALYs.
Source of data for measurement of HRQL	Yes	Measurements were taken from patients with bipolar disorder.
Source of preference data for valuation of changes in HRQL	Yes	EuroQol 5-Dimension (EQ-5D) was used to measure health-related quality of life.
Discount rate	Yes	3.5% per annum for costs and QALYs
Equity weighting	Yes	All QALYs gains were treated equally.

5.2.2 Model structure.

The manufacturers provided a *de novo* model-based economic evaluation constructed in Microsoft Excel© and based upon a cohort Markov model. In addition to an absorbing death state, the manufacturers modelled 22 different health states, divided into four distinct groups. Three of these groups related to antipsychotic treatment lines (first-, second- and third-line) and were identical in structure: each contained an acute phase (consisting of three separate health states based on elapsed time); a sub-acute phase; and a maintenance phase (consisting of two separate health states based on whether or not the patient was assumed to be on treatment). The fourth group consisted of four separate health states which modelled therapy resistance for patents who had not responded to the three lines of antipsychotic treatment.

A schematic of the model is shown in Figure 2 (Figure B16; page 147 of MS). The modelling of adverse events was included within the treatment-related health states. Patients were modelled as receiving in-hospital treatment for all of the health states within the acute and sub-acute phases, as well as the "Therapy Resistance Hospitalised" state. Patients could die at any point in the model.

Figure 2: A schematic of the manufacturers' model (from Figure B16; page 147 of MS)



All patients enter the model at the start of the first treatment line (in the health state "Acute 1st line Week 1"). Patients move through the treatment lines if either they discontinue drug use before response (i.e. during the acute phase) or if they relapse before discharge from hospital (i.e. during the sub-acute phase). If patients relapse within the maintenance phase they remain on the same treatment line to which they responded. The "Therapy Resistance Phase" is essentially the fourth and final treatment line, where treatment is assumed to be lithium. The therapy resistance phase is discussed in Section 5.2.7. QALYs are accrued as time spent in each health state, with different utility values for the acute, sub-acute and maintenance phases (the lowest utility values are for the acute phase and the highest are for the maintenance phase). The main driver of costs is time spent as an inpatient (in either the acute or sub-acute phases). Treatment effectiveness was reflected by the time spent within each model phase (acute, sub-acute or maintenance).

The Markov model used a cycle length of one week, to reflect the timing of assessments in the pivotal trial for aripiprazole.[21] The ERG believes that weekly Markov cycles are appropriate because the first three weeks of acute treatment are the main drivers of cost-effectiveness results. The option for half-cycle correction was included in the model but was not used in the base case results presented in the MS. Including a half-cycle correction slightly reduces the total costs and total QALYs for each strategy, but does not alter the conclusions of the economic evaluation.

5.2.3 Interventions, comparator and treatment sequences.

The manufacturers considered four different antipsychotics in the treatment sequence; aripiprazole, risperidone, quetiapine and olanzapine. The daily costs of these drugs, as used in the economic evaluation, are presented in Table 38. Non-proprietary costs were used for each of the antipsychotics, with the exception of aripiprazole, for which the branded cost (Abilify®) was used as this drug is still under patent. The manufacturers note that a generic version of their drug is expected in 2014 (page 217 of MS).

Table 51: Daily antipsychotic drug costs used in the model

Antipsychotic	Daily cost (£)	Cost of 12-week treatment course (£)
Aripiprazole	5.13	430.92
Risperidone	0.06	5.04
Quetiapine	4.04	339.36
Olanzapine	1.32	110.88

The manufacturers defined the treatment sequence of risperidone, quetiapine and olanzapine to represent usual care, in the MS this was labelled as 'Strategy 1'. For their base case analyses the manufacturers assumed that olanzapine would be replaced with aripiprazole; the position of aripiprazole in the treatment sequences was varied, giving different strategies. Use of quetiapine was restricted to always occur after use of risperidone, resulting in the following possible treatment sequences (denoted with the strategy name used in the MS):

Strategy 1: (S1) (base case)**:** risperidone, quetiapine, olanzapine.
Strategy 2: (S2) risperidone, aripiprazole, quetiapine.
Strategy 3: (S3) aripiprazole, risperidone, quetiapine.
Strategy 4: (S4) risperidone, quetiapine, aripiprazole

In a scenario analysis the manufacturers considered replacing quetiapine with olanzapine (in Strategy 1 their positions are reversed).

The manufacturers' base case sequence was based on clinical opinion; although their clinical advisors noted that treatment sequences are often tailored to the individual patient's needs (see below). The choice of which drug should be replaced by aripiprazole (quetiapine or olanzapine) was somewhat arbitrary, although both choices are considered in the MS. Both the manufacturers and the ERG separately sought clinical advice about current treatment sequences and the role of aripiprazole in this. There was general consensus around the following points:

- The choice of treatment sequence is usually dependent upon severity of symptoms, side-effect profile, comorbidities and the likelihood of adherence.

- Aripiprazole is already in routine use.
- Any of the antipsychotics may be considered as first-line treatment.

The ERG notes that, based on the data used in the manufacturers' economic evaluation, risperidone has the highest probability of YMRS response at week 3, and it is substantially cheaper than the other antipsychotics. Because of this, the use of risperidone is first line in the manufacturers' base case, and the ERG considers the constraint to always use risperidone before either quetiapine or olanzapine to be reasonable, although the ERG notes that the actual treatment sequence offered is likely to depend upon patient characteristics.

Based on the data used by the manufacturers, the use of olanzapine may be preferable to use of quetiapine, as the former has a lower probability of discontinuation, higher probability of YRMS response (both at week 3) and is cheaper. This is reflected by the results reported in the MS; Strategy 1 of the scenario analysis (where olanzapine is sequenced before quetiapine) dominates the base case. Further details are provided in Table 52.

Table 52: Comparison of treatment sequences without aripiprazole

Sequence	Total costs (£)	Total QALYs	Reference in MS
RIS, QUE, OLA	75,066	2.51637	Table 2, p15
RIS, OLA, QUE	74,687	2.51672	Table B87, p238

RIS: risperidone, QUE: quetiapine, OLA: olanzapine

Therefore the ERG believe that a sequence of risperidone; olanzapine; quetiapine should also be considered. Furthermore, the ERG believe that since all four of the antipsychotics are currently in use, an appropriate intervention-sequence is represented by the use of all four antipsychotics. An analysis considering all four antipsychotics was requested by the ERG from the manufacturers (question B17, page 36 of clarification response), but this was not performed due to time constraints.

5.2.4 Time horizon, discounting and length of treatment.

All patients enter the model at the age of 15. The time horizon employed in the model was until patients reach adulthood (at the age of 18), at which point treatment management options change, resulting in a time horizon of three years. Although many technologies have impacts over a patient's lifetime, the ERG considers that the use of a restricted time horizon in this submission is justifiable, due to the following reasons;

i. NICE guidance recommends that treatment management options change at the age of 18.
ii. There is no modelled difference in mortality rates for any of the strategies.
iii. The main driver of cost-effectiveness relates to the acute phase of inpatient care.

However, the ERG notes that the choice of starting age was somewhat arbitrary, and requested that the manufacturers explore the use of different starting ages. Results are reproduced in Section 5.2.11.2 and show that the conclusions of the manufacturers' base case analysis remain unaltered irrespective of the selected starting age.

Both costs and QALYs are discounted at an annual rate of 3.5%, as recommended in the NICE reference case.[65] In the manufacturers' model this is achieved by using the standard discount factor of $1/(1+0.035)^t$, treating t as an integer. Hence for the first year (52 cycles) a value of $t = 0$ is used, for the second year a value of $t = 1$ is used, and so-on. The ERG considers that it is more appropriate to treat t as a fraction (setting it equal to 'cycle number / 52'). However this does not have a material impact upon the results.

The manufacturers note that positive CHMP opinion was granted for up to 12 weeks of treatment. This 12-week limit is reflected in the model structure; within each treatment-line patients spend between 6 and 8 weeks receiving treatment in hospital, with treatment post-discharge being reported to be an average duration of 4 weeks. However, the ERG notes that:

- Within the model it is possible for patients to receive treatment post-discharge for longer than 4 weeks, with the result that some patients receive more than 12 weeks of treatment.
- The ERG's clinical advisors stated that the average length of treatment with each of the antipsychotics was normally far in excess of 12 weeks, being closer to 12 months and usually at least six months.

The ERG's clinical advisors also questioned the length of treatment during the acute phase. In the manufacturers' model the acute phase lasted up to 3 weeks; with patients able to respond or fail after one, two or three weeks. It was stated that the acute phase normally exceeds 3 weeks because it usually takes at least one month to assess if a patient has responded or failed to respond to antipsychotic treatment.

During the clarification process, the ERG requested additional analyses from the manufacturers to explore the impact of changing treatment duration during both the acute and euthymic: treated phases of the model. Results are presented in Section 5.2.11.2 of this report. In addition, the ERG amended the manufacturers' model to limit antipsychotic treatment to a maximum of twelve weeks, results of this amendment are presented in Section 5.3 of this report. For all of the analyses considered Strategy two (aripiprazole second-line) dominated each of the other strategies.

5.2.5 Population

The MS (Section 6.2.1) states that the patient population used in the economic evaluation reflects the population detailed in the CHMP indication: namely patients aged between 13 and 17 years with manic

episodes of bipolar I disorder. However, the effectiveness data for aripiprazole used in the model uses the entire population from its pivotal trial[21] which relates to patients aged between 10 and 17 years with manic or mixed episodes of bipolar I disorder. The NICE scope states the intended population to be "children and adolescents with acute manic or mixed episodes associated with bipolar I disorder". The effect of patient starting age is discussed in Section 5.2.4 of this report. The manufacturers were asked to demonstrate the applicability of their cost-effectiveness results to both manic and mixed populations. The ERG is satisfied with their response (question B1, pages 23-24 of clarification response), which showed both that the data used contained a mixture of manic and mixed patients, and that there are no data to suggest a differential response to aripiprazole between these two populations.

The ERG notes that depression health states were included in the manufacturers' conceptual models (see Appendix 14 Figure 2 of the MS), but were not included in the final Markov model. The manufacturers justify this (both in the original submission and in the clarification response) on the basis of lack of data and the fact that the submission is related to the treatment of manic episodes.

5.2.6 Treatment effectiveness, mortality and adverse events.

Treatment effectiveness is measured by the probability of both discontinuation and of response. Due to a lack of data, differential effectiveness between the drugs is only modelled during the acute (three-week) phase. Data for the acute phase comes from the results of a network meta-analysis conducted by the manufacturer. The appropriateness of this analysis is discussed in Section 4.4 of this report. Data beyond the acute phase was based on expert opinion. The effectiveness data used by the manufacturers are summarised in Table 53. The manufacturers assumed that the same effectiveness data applied for each treatment line.

Table 53: Effectiveness data used in the manufacturers' base case

Probabilities	**aripiprazole**	**risperidone**	**quetiapine**	**olanzapine**
Discontinuation – data from network meta-analysis ɛ				
0 to 1 week	2.06%	2.48%	1.97%	1.54%
1 to 2 weeks	8.56%	10.30%	8.17%	6.37%
2 to 3 weeks	7.00%	8.43%	6.70%	5.23%
0 to 3 weeks	**17.62%**	**21.21%**	**16.84%**	**13.14%**
YMRS Response – data from network meta-analysis ɛ				
0 to 1 week	28.09%	23.89%	22.20%	20.92%
1 to 2 weeks	22.59%	26.48%	17.86%	30.82%
2 to 3 weeks	8.40%	9.70%	6.64%	0.60%
0 to 3 weeks	**59.08%**	**60.07%**	**46.70%**	**52.34%**
Weekly probability of relapse (in the sub-acute and maintenance phases) – based on expert opinion.				
Whilst treated*	0.57% for all drugs			
Not treated	0.67% for all drugs			

* In the model this corresponds to both the 'Responder observation' and 'Euthymic treated' health states.

ɛ Numbers represent the proportion of the starting cohort. Thus summation of the three individual values equal the 0-3 week values

The ERG is satisfied with the model results being driven by effectiveness during the acute phase. However, the ERG's clinical advisors felt that the post-acute relapse rates were too low. The rate of relapse of 0.57% per week whilst treated was based on expert opinion to the manufacturers which assumed the rate was 5% over the entire duration of treatment. The ERG requested that the manufacturers ran analyses using higher rates; details of these analyses are presented in Section 5.2.11.2 and showed that whilst increasing relapse rates led to higher accumulated costs and lower accumulated QALYs for all the strategies, the incremental analyses remained the same.

The manufacturers' model assumes that the effectiveness of each antipsychotic is independent of which treatment line it is used in. The ERG notes that effectiveness data come either from trials in which the antipsychotics were used as first-line treatment, or are based on expert opinion. Furthermore, the patient populations entering each treatment line are likely to have different characteristics. For example, patients entering third line treatment represent the subset of patients who had failed to respond to two different antipsychotics – these failures may represent an inherent resistance to antipsychotic drug therapy. Whilst there were no data on how effectiveness reduces across treatment lines, this could have been explored in a sensitivity analysis by assuming an arbitrary reduction in the efficacy of the antipsychotics with each additional line of treatment. Through the clarification process the ERG requested that the manufacturers conduct such a sensitivity analysis. The resulting analyses are discussed in Section 5.2.11.2, for all of the reductions tested by the manufacturers' use of aripiprazole second line (S2) dominated each of the other treatment strategies.

It is also noted that neither combination therapy with mood stabilisers nor the inclusion of psychological therapy (such as cognitive-based therapy) as adjunctive treatment are considered in the model. However the ERG, in consultation with their clinical advisors, acknowledges that there are little data, either on how their inclusion may alter effectiveness or on how their effectiveness may differ between the antipsychotics. Given this, the ERG believes that the inclusion of either of these would not substantially alter the overall conclusions.

For their base case analysis, the manufacturers pooled efficacy and safety data for two separate doses of aripiprazole; 10mg and 30mg (both daily). In their submission the manufacturers stated that they expected the licence for aripiprazole to be at 10mg per day (MS page 155). The impact on the cost-effectiveness results of using 10mg per day was explored in a scenario analysis and was found to have no substantial impact; results are presented in Table 64 (Section 5.2.11.2) of this report. In addition, the ERG's clinical advisors stated that the modelling of pooled doses may be more likely to reflect current practice.

Due to a lack of data, the manufacturers assumed that the rate of mortality was the same for all of the antipsychotics. Mortality rates were based on UK life-tables (broken-down by age and gender)[66] and increased to reflect the higher rates of mortality observed amongst patients with bipolar disorder. The mortality rate ratios employed in the model are 10.09 for males and 24.93 for females.[67] The manufacturers make a slight mistake in applying these multipliers directly to the general population mortality probabilities instead of applying them to the general population mortality rates. However, this makes a negligible difference to the results.

The manufacturers included three adverse events within their model: EPS; somnolence; and weight gain. The incidence of these events was taken from the manufacturers' network meta-analysis. Data were missing on EPS and somnolence for olanzapine, with the incidence of these set equal to the lowest incidence of the other antipsychotics. The network meta-analysis also considered prolactin increase. However, this was not included in the manufacturers' model due to a lack of data on related side-effects. It is noted that these two decisions (the imputation for olanzapine and not modelling prolactin increase) will both create a bias in the cost-effectiveness results, which will disfavour aripiprazole. An overview of the role of each adverse event in the manufacturers' model is presented in Table 54.

Table 54: Overview of the adverse events considered in the economic evaluation

	EPS	Somnolence	Weight gain	Prolactin increase
Median incidence (during the acute phase), from network meta-analysis				
aripiprazole	0.158	0.463	0.034	0.002
risperidone	0.112	0.266	0.041	0.321
quetiapine	0.116	0.273	0.403	0.070
olanzapine*	0.112	0.266	0.450	0.416
Utilities; affected health states	Acute phase	Acute phase	Acute phase and Euthymic treated	None
Utility multiplier	0.722	0.905	0.865	N/A
Costs; affected health states	None	None	Euthymic treated	None
Weekly cost	N/A	N/A	£16.57	N/A

* Values for EPS and somnolence set equal to the lowest value for the other antipsychotics.

The weekly cost for weight gain is applied until the end of the model.

The choice of adverse events considered in the manufacturers' economic evaluation was based on a mixture of available data and the existing literature (including clinical guidelines). The ERG notes that cardiovascular events were not considered, even though clinical advisors to the ERG felt that this could be an issue.

The manufacturers did not directly include costs relating to any of the adverse events during the acute phase, stating that these costs would be indirectly included via their use of NHS Reference Costs for this time period. However, the ERG considers that this approach would not reflect the differential costs associated with the use of each antipsychotic due to differential rates of adverse events. The ERG requested that the manufacturers explicitly modelled the cost of drug-related adverse events during the inpatient period. Details of this analysis are presented in Section 5.2.11. Due to uncertainty in resource use two different scenarios were tested. For both of these, use of aripiprazole second line (S2) dominated each of the other treatment strategies.

Of the adverse events presented, only the effects of weight gain are modelled beyond the acute phase. The ERG notes that this approach will favour aripiprazole as it has the lowest incidence of weight gain of the four antipsychotics. However, the ERG's clinical advisors believe that this is the only adverse event that would have an impact beyond the acute phase, so the manufacturers' approach appears reasonable.

5.2.7 Therapy resistance

Within the model, patients who do not respond to three lines of antipsychotic treatment enter the therapy resistance treatment line. Therapy-resistant patients are assumed to receive five weeks of inpatient treatment with lithium before being discharged (the model assumes that lithium has a 0% probability of response). Following discharge, patients remain on lithium treatment, and have a weekly probability of (spontaneous) response of 1.07%. If patients respond, they are modelled in the same manner as if they responded from an antipsychotic drug, with the exception that if they relapse they return to inpatient treatment with lithium.

The ERG is satisfied with the modelling of therapy resistance, with the caveat that, following clinical advice, the ERG believe that four lines of antipsychotics would usually be tried before patients are deemed to be therapy resistant.

5.2.8 Health related quality of life

Within the pivotal trial NCT00110461[21] for aripiprazole HRQoL was measured using the Pediatric Quality of Life Enjoyment and Satisfaction Questionnaire. Because this is not a preference-based measure of HRQoL it was not used in the manufacturers' model. The manufacturers conducted a systematic review to identify preference-based utility values for paediatric bipolar disorder, but were unable to identify any articles. In addition the manufacturers were unable to identify any reliable methods for mapping from the non-preference-based method to preference-based methods. The ERG does not believe that the manufacturers have omitted any relevant studies or mapping methods.

To populate utility values in their economic model, the manufacturers used the results of studies looking at HRQoL in adult populations with bipolar disorder (identified from their systematic review). Two studies were identified: one study presented EQ-5D data from a UK population[68] and was used in the manufacturers' main analysis; the other study presented EQ-5D data from a USA population[69] and was used in a sensitivity analysis. Utilities for the included adverse events were taken from separate studies: those for weight gain were based on the general population;[70] whilst those for somnolence and EPS came from patients with schizophrenia.[71,72]

The manufacturers acknowledge that using utility data from an adult population is a limitation, but that it is justified by the lack of relevant data for the paediatric population. To take into account differences in utility by age, the manufacturers converted the reported utility values into multiplicative decrements (relative to the age-gender matched adult general population). These were then applied to utility values for the age-gender matched paediatric general population, which were calculated from a published formula.[73]

Table 55: Utility values used by the manufacturers in their economic evaluation

Utility values	Main analysis	Sensitivity analysis
General population	= 0.951+0.021*% Male – Age * $2.59*10^{-4}$ - Age^2 * $3.32*10^{-5}$ Examples (54% male); 15 year old: 0.951 16 year old: 0.950 17 year old: 0.948	-
Multipliers		
Acute phase	0.775	0.259
Responder observation	0.954	0.849
Euthymic treated	0.954	0.933
Euthymic not treated	0.954	0.832
Therapy resistant inpatient	0.809	0.292
Therapy resistant outpatient	0.809	0.674
EPS	0.815	-
Weight gain	0.908	0.926
Somnolence	0.905	-
Decrement		
Hospitalisation (decrement)	0.070	0 (assumed to be included in above values)

In the model general population utility values are updated every cycle.

The ERG notes that the formula used to calculate utility values for the general paediatric population was developed in a general adult population. The ERG also notes that EQ-5D may not be the most appropriate preference-base measure for use in children.[74] However, as with the remainder of the utility values, the use of these appears to be reasonable given the lack of relevant data.

There are some large discrepancies in the bipolar-related utilities reported in the two studies used by the manufacturers. For example, the utility multiplier for the acute phase is 0.775 in the main analysis, but 0.259 in the sensitivity analysis. A similar difference is seen for therapy resistant inpatients, with multipliers of 0.809 and 0.292. However, use of the alternative utility multipliers does not alter the incremental cost-effectiveness results, with use of aripiprazole second-line (S2) dominating each of the other treatment strategies.

5.2.9 Resources and costs

Daily in-hospital costs were based on NHS Reference Costs 2010/11[75] (code MHIPC1; NHS Trusts Mental Health Inpatients – Children). In the MS this cost was assumed to include costs relating to adverse events, but not the cost of the antipsychotic drugs. This assumption is discussed in more detail in Section 5.2.6 of this report.

Out-of-hospital resource use was based on clinical opinion to the manufacturers, with costs taken from the PSSRU.[76] Drug costs (where appropriate) were included separately, as were costs related to weight-gain. An overview of the costs used in the model is presented in Table 56.

Table 56: Weekly costs (£) used within the model

Inpatient care (acute and therapy resistant phases)*	4 214.68
Outpatient care; on treatment*	80.37
Outpatient care; no treatment	42.33
Outpatient care; therapy resistant*	161.84
Weight gain	16.57
Aripiprazole	35.90
Risperidone	0.40
Quetiapine	28.31
Olanzapine	9.24
Lithium	0.25

* Excludes drug costs

5.2.10 Base case cost effectiveness results

The manufacturers' base case results are shown in an incremental analysis (in ascending order of cost) in Table 57 (reproduced from Table B78; MS page 229). It should be noted that the manufacturers' base case results did not change following the ERG's clarification questions.

Table 57: Base case results reported in the manufacturers' submission

Sequence	Total costs (£)	Total QALYs	Incremental costs (£)	Incremental QALYs	ICER
S2 (RIS, ARI, QUE)	£74,133	2.52466			
S3 (ARI, RIS, QUE)	£74,379	2.52348	£246	-0.0012	Dominated by S2
S4 (RIS, QUE, ARI)	£74,888	2.52297	£755	-0.0017	Dominated by S2
S1 (RIS, QUE, OLA)	£75,066	2.51637	£933	-0.0083	Dominated by S2

RIS: risperidone. QUE: quetiapine. OLA: olanzapine. ARI: aripiprazole

The manufacturers noted that, based on the cost-effectiveness results, use of aripiprazole at any point in the treatment pathway (S2, S3 or S4) dominates a sequence in which the drug is not included (S1). The manufacturers' additional sensitivity analyses explored the robustness of this conclusion. The manufacturers did not declare that the use of aripiprazole second-line (after risperidone first-line) dominates all of the other treatment strategies considered, when using mean cost-effectiveness results. However, the ERG has been advised that all four antipsychotics are typically tried in patients before they are declared treatment resistant. Hence a treatment strategy incorporating all four drugs should be considered. As previously mentioned, the

ERG requested that the manufacturers conduct such an analysis, but they were unable to do so within the available time-frame.

The MS provided a breakdown of results (undiscounted costs and QALYs) for each strategy and each of the 23 health states (Tables B71 to B76; MS pages 220-228). This breakdown is summarised in Table 58 and Table 59, which group the health states into fewer categories.

Table 58: Undiscounted costs by broad health state for each strategy

Undiscounted costs	Strategy One (RIS, QUE, OLA)	Strategy Two (RIS, ARI, QUE)	Strategy Three (ARI, RIS, QUE)	Strategy Four (RIS, QUE, ARI)
Acute	£14,714	£14,714	£14,675	£14,714
Resp. Obs.	£20,494	£20,494	£20,148	£20,494
Euthymic*	£4,459	£4,459	£4,546	£4,459
First line	**£39,666**	**£39,666**	**£39,369**	**£39,666**
Acute	£9,486	£9,760	£9,910	£9,486
Resp. Obs.	£9,569	£13,363	£13,764	£9,569
Euthymic*	£1,986	£2,724	£2,700	£1,986
Second line	**£21,040**	**£25,846**	**£26,374**	**£21,040**
Acute	£7,144	£6,154	£6,154	£7,282
Resp. Obs.	£8,345	£6,193	£6,193	£9,952
Euthymic*	£1,572	£1,179	£1,179	£1,903
Third line	**£17,062**	**£13,526**	**£13,526**	**£19,137**
Therapy resistant	**£25,266**	**£22,920**	**£22,920**	**£22,920**
Total	***£103,034***	***£101,958***	***£102,189***	***£102,763***

Resp. Obs.: Responder observation
* Includes 'Euthymic treated' and 'Euthymic not treated'

Table 59: Undiscounted QALYs by broad health state for each strategy

Undiscounted costs	Strategy One (RIS, QUE, OLA)	Strategy Two (RIS, ARI, QUE)	Strategy Three (ARI, RIS, QUE)	Strategy Four (RIS, QUE, ARI)
Acute	2.3	2.3	2.1	2.3
Resp. Obs.	3.9	3.9	3.7	3.9
Euthymic*	90.8	90.8	88.8	90.8
First line	**96.9**	**96.9**	**94.6**	**96.9**
Acute	1.4	1.4	1.5	1.4
Resp. Obs.	1.7	2.4	2.6	1.7
Euthymic*	38.6	52.7	54.7	38.6
Second line	**41.8**	**56.5**	**58.8**	**41.8**
Acute	1.0	0.9	0.9	1.1
Resp. Obs.	1.5	1.1	1.1	1.8
Euthymic*	31.1	22.7	22.7	36.6
Third line	**33.7**	**24.8**	**24.8**	**39.4**
Therapy resistant	**52.6**	**47.4**	**47.4**	**47.4**
Total	***224.97***	***225.64***	***225.58***	***225.55***

Resp. Obs.: Responder observation
* Includes 'Euthymic treated' and 'Euthymic not treated'

The results presented in Table 58 and Table 59 show that when aripiprazole replaces olanzapine, it results in higher costs and higher QALYs for the treatment line during which it is used (these additional costs and QALYs stem from the 'Responder observation' and 'Euthymic' health states), but results in lower costs and lower QALYs for all subsequent lines of treatment (including the fourth line of therapy resistance). The initial increase in costs is outweighed by the subsequent decrease, whereas the initial increase in QALYs outweighs the subsequent decrease, with the result that the inclusion of aripiprazole at any point (S2, S3 and S4) dominates its exclusion (S1).

5.2.11 Sensitivity analyses

5.2.11.1 Probabilistic sensitivity analysis (PSA)

For the manufacturer's PSA, values for the following variables were taken from their posterior distributions, as generated from the network meta-analysis:

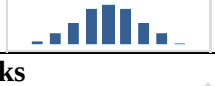
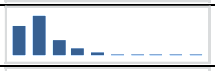
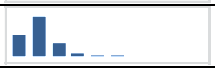
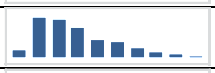
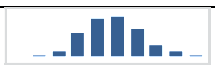
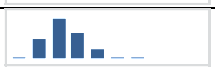
- Probability of YMRS response at three weeks
- Relative risk (compared with placebo) of discontinuation at three weeks
- Incidence of EPS (during the acute phase)
- Incidence of clinically significant weight grain (during the acute phase)
- Incidence of somnolence (during the acute phase)

The manufacturers noted that if utility values were independently sampled in their model then potentially illogical values could result; for example patients with acute mania could be modelled as having a better quality of life than patients who had responded to treatment. Instead, the manufacturers sampled utility values for patients who had responded to treatment, multiplying by 0.85 to give utility values for treatment resistant patients and by 0.81 to give utility values for patients with acute mania.

Whilst the ERG considers that the monotonicity between the utility health states should be maintained, it is noted that the manufacturers have ignored uncertainty in their multipliers and hence also in their assessment of uncertainty in the utility values.

An overview of the key inputs to the manufacturers' PSA is presented in Table 57. The PSA inputs for these parameters all came from the manufacturers' network meta-analysis (see Section 4.4 of this report).

Table 60: Details of the inputs to the manufacturers' PSA

	Mean	Minimum	Maximum	Graph
Probability of YMRS response at three weeks				
aripiprazole	0.59	0.40	0.80	
risperidone	0.60	0.38	0.79	
quetiapine	0.47	0.27	0.65	
olanzapine	0.52	0.31	0.75	
Relative risk (compared with placebo) of discontinuation at three weeks				
aripiprazole	0.75	0.31	1.58	
risperidone	0.92	0.33	2.64	
quetiapine	0.71	0.27	1.76	
olanzapine	0.57	0.16	1.39	
Incidence of EPS (during the acute phase)				
aripiprazole	0.17	0.02	0.50	
risperidone	0.12	0.01	0.48	
quetiapine	0.15	0.01	0.79	
olanzapine	0.12	0.01	0.48	
Incidence of weight grain (during the acute phase)				
aripiprazole	0.04	0.00	0.17	
risperidone	0.05	0.00	0.30	
quetiapine	0.39	0.03	0.98	
olanzapine	0.53	0.06	1.00	
Incidence of somnolence (during the acute phase)				
aripiprazole	0.47	0.18	0.85	
risperidone	0.27	0.10	0.53	
quetiapine	0.27	0.09	0.55	
olanzapine	0.27	0.10	0.53	

Std. Dev.: Standard deviation YMRS: Young Mania Rating Scale

Within each Section graphs are plotted on the same horizontal axis.

The ERG notes that within the manufacturers' PSA, probabilities of YRMS response and discontinuation were modelled separately. This meant that it was possible for their combined probability of occurrence could exceed 100%. However, using week-three values, this only occurred for aripiprazole in 4 (0.4%) of the PSA

samples and for risperidone in 29 (2.9%) of the PSA samples (it did not occur for quetiapine or olanzapine). The ERG corrected this by reducing both of the probabilities (by the same amount) whenever a violation occurred so that their sum would equal 100%. This had a minor effect on the PSA results, increasing costs by about 0.03% and reducing QALYs by about 0.005% for each of the strategies. Results of the incremental analyses remained unchanged, with use of aripiprazole second-line (S2) dominating each of the other treatment strategies.

To use their network meta-analysis results in their PSA the manufacturers used a random sample of 1,000 from 50,000 iterations of their fixed-effects model. The number of iterations required for convergence of their fixed-effects model was based on graphical methods (from questions A14 and A15, page 16 of clarification response). However, the manufacturers did not provide details of how they assessed the number of PSA runs required for stable results. The ERG checked the PSA runs graphically for convergence, the results are displayed in Figure 3 and Figure 4 and suggest adequate convergence of the PSA runs.

Figure 3: Cumulative average cost for each strategy against PSA run

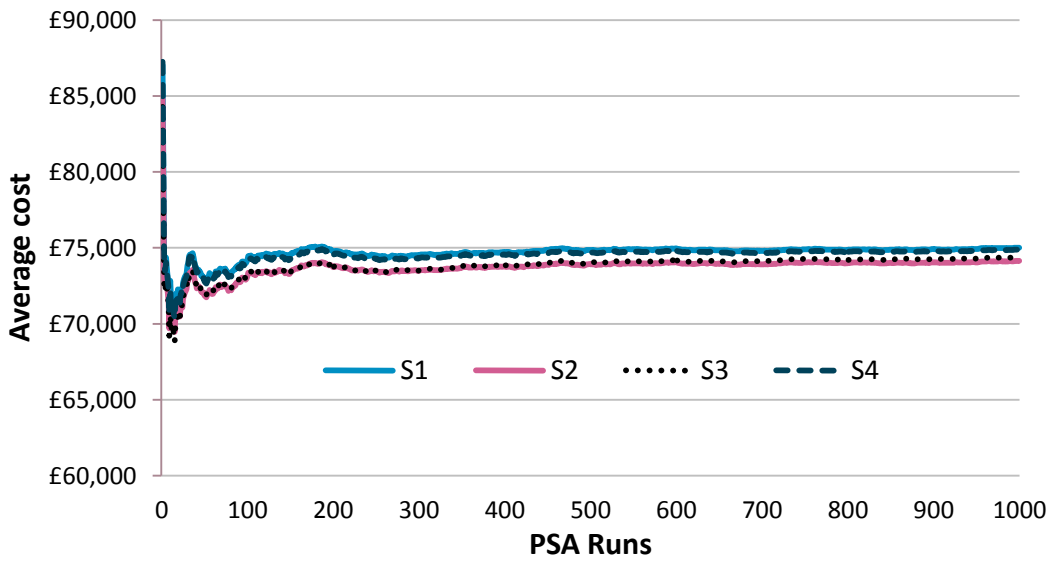
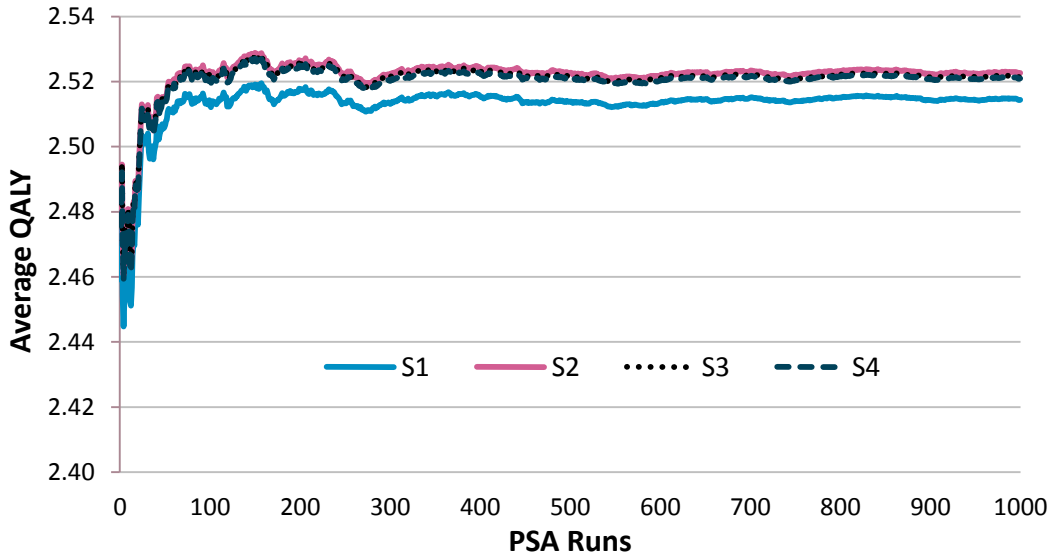


Figure 4: Cumulative average QALYs for each strategy against PSA run



In the MS the results of their PSA were presented as scatter plots of incremental costs and QALYs for scenarios two to four (all relative to scenario one), these are reproduced in Figures 4 to 7 (Figures B19 to B21; MS pages 231/232).

Figure 5: Scatter-plot of incremental costs and incremental QALYs; strategy 2 vs. strategy 1

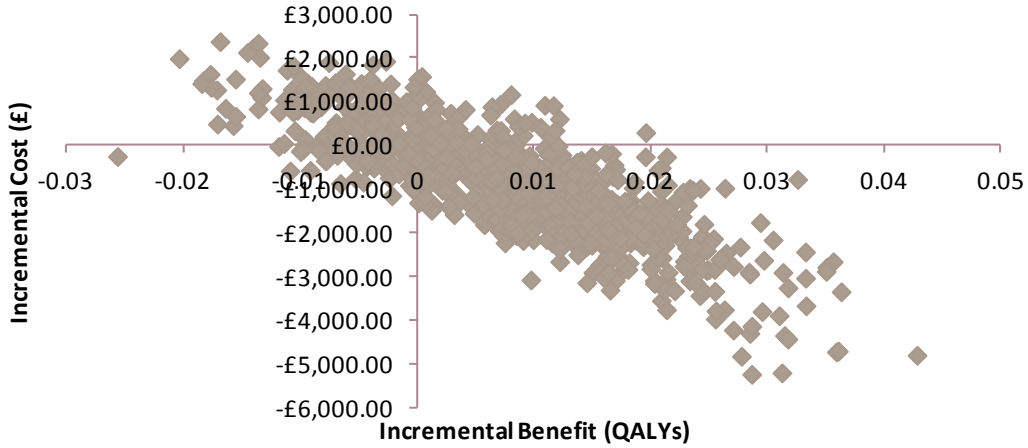


Figure 6: Scatter-plot of incremental costs and incremental QALYs; strategy 3 vs. strategy 1

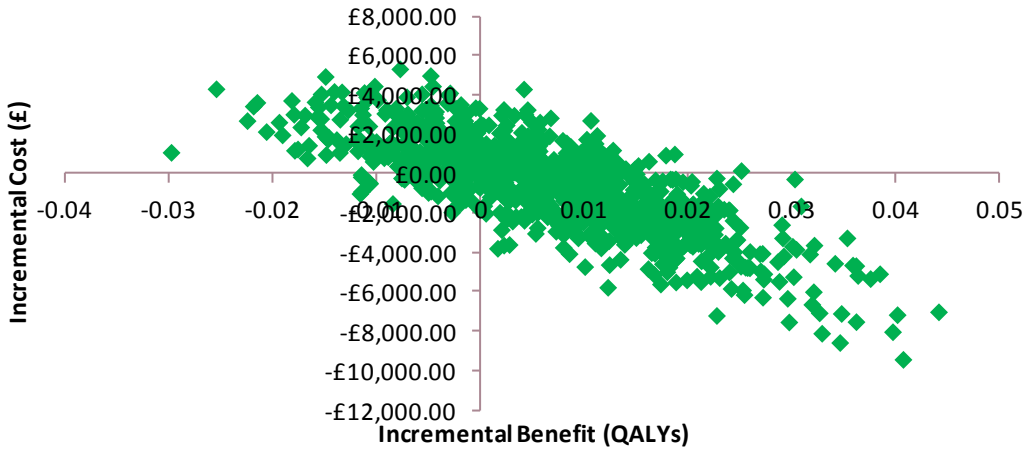
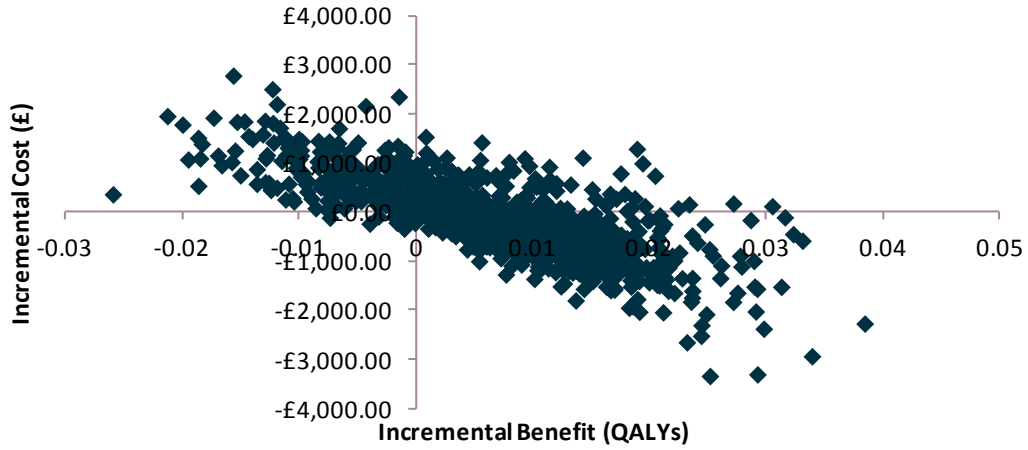


Figure 7: Scatter-plot of incremental costs and incremental QALYs; strategy 4 vs. strategy 1



Figures 5 to 7 show considerable uncertainty in the cost-effectiveness results. For each of the three strategies incorporating aripiprazole (S2, S3 and S4) the majority of the PSA results fall in the bottom-right quadrant, indicating dominance over the strategy excluding aripiprazole (S1).

No summary results were reported in the MS. However, based on the results presented in the manufacturers' model the ERG generated Table 61.

Table 61: Results from the manufacturers' PSA

Costs	Deterministic mean	PSA mean	PSA 95% interval*
S1	£75,066	£75,016	£61,207 to £90,979
S2	£74,133	£74,138	£60,003 to £90,969
S3	£74,379	£74,386	£59,663 to £90,343
S4	£74,888	£74,884	£61,066 to £91,076
QALYs			
S1	2.516	2.514	2.336 to 2.649
S2	2.525	2.523	2.343 to 2.656
S3	2.523	2.521	2.341 to 2.655
S4	2.523	2.521	2.343 to 2.654
ICER	**Based on incremental analysis**		
S2	-	-	-
S3	Dominated by S2		Dominating S2 to dominated by S2
S4	Dominated by S2		Dominating S2 to dominated by S2
S1	Dominated by S2		Dominating S2 to dominated by S2

* Percentile-based.

Table 61 shows that there is little difference between the mean deterministic and PSA results: mean PSA costs for S1 reduce by £50; the costs for the other strategies change by less than £7. All of the mean PSA QALYs are reduced relative to the deterministic means, but this reduction is always less than 0.0021 QALYs. All of the changes (for both costs and QALYs) are less than 0.1% of the deterministic value.

A comparison of the ranges for costs and QALYs across the strategies shows how similar they are. However, a more insightful analysis is to compare incremental costs and QALYs simultaneously through calculation of ICERs. This analysis is presented at the bottom of Table 61 and shows that, whilst the mean results from both the deterministic analysis and PSA indicate that S2 dominates all of the other treatment strategies, a 95% confidence interval about this result also includes the possibility of each treatment strategy dominating S2.

Scatter plots comparing incremental costs and QALYs for each strategy compared with S2 are displayed in Figures 8 to 10. These graphs have been fixed so that they have common scales. Again the uncertainty in the cost-effectiveness results is emphasised. Figures 9 and 10 also show that there are strong correlations amongst the outcomes of the three treatment strategies containing aripiprazole (S2, S3 and S4). Comparisons of every strategy against every other strategy are summarised in Table 62 which shows the percentage of times each strategy either dominates, or is dominated by every other strategy. Figure 11 displays the cost-effectiveness acceptability curve (CEAC), which shows the probability of each strategy being the most cost-effective for willingness-to-pay thresholds between £0 and £100,000. The CEAC may appear to be counter-intuitive, as S3 has a higher probability of being cost-effective than S2, even though S2 dominates S3. Similarly S1 has a

higher probability of being cost-effective than S4, even though S4 dominates S1. These results are due to both the correlated nature of the different treatment strategies, allied with the fact that the CEAC does not take into account the consequences of not being cost-effective.

Figure 8: Scatter-plot of incremental costs and incremental QALYs: strategy 1 vs. strategy 2

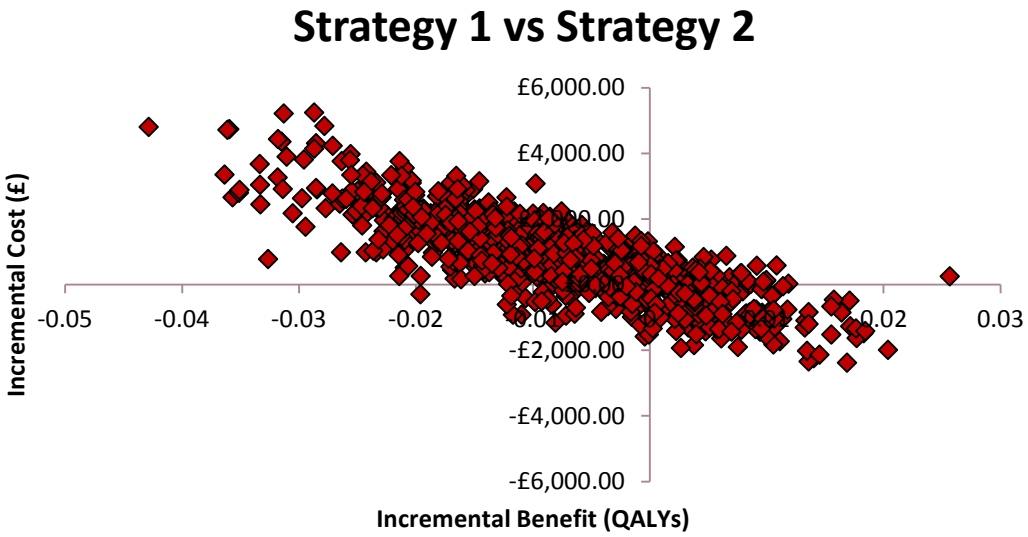


Figure 9: Scatter-plot of incremental costs and incremental QALYs: strategy 3 vs. strategy 2

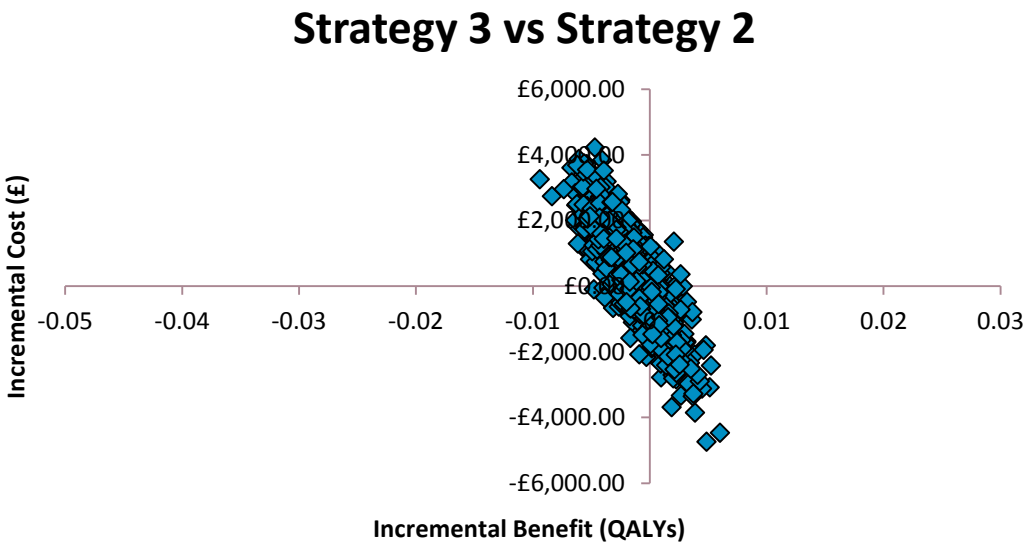


Figure 10: Scatter-plot of incremental costs and incremental QALYs: strategy 4 vs. strategy 2

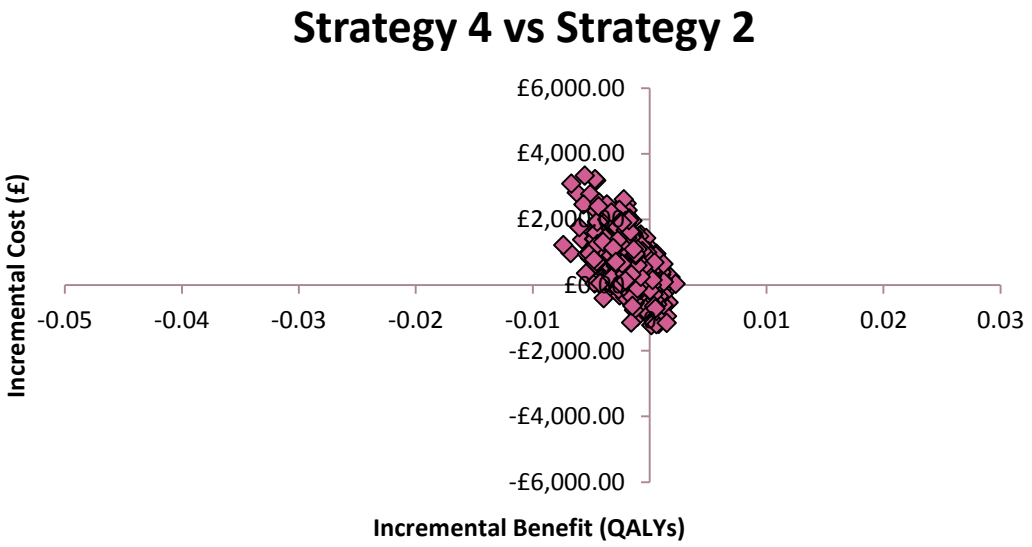


Figure 11: Cost-effectiveness acceptability curve for the four treatment strategies

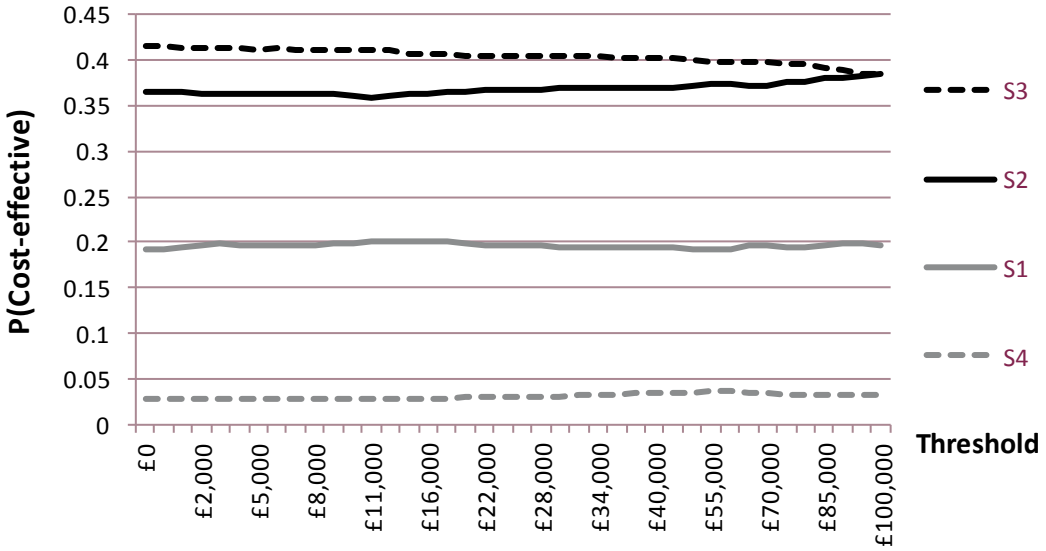


Table 62: Summary of PSA Results: dominance compared with strategies 1 and 2

	Results compared with S1		Results compared with S2		Results compared with S3		Results compared with S4	
Scenario	% Dominating	% Dominated	% Dominating	% Dominated	% Dominating	% Dominated	% Dominating	% Dominated
S1	-	-	14.3	72.1	22.5	54.4	22.9	57.2
S2	72.1	14.3	-	-	52.8	27.0	84.0	4.5
S3	54.4	22.5	27.0	52.8	-	-	48.6	31.8
S4	57.2	22.9	4.5	84.0	31.8	48.6	-	-

The results from the manufacturers' PSA indicate that, if a three-drug treatment sequence is used, aripiprazole should be included, although its position within the treatment sequence remains unclear. The strategy that excludes aripiprazole (S1) is dominated by each of the other strategies in over half of the PSA results. Furthermore, the probability that S1 is the most cost-effective strategy is roughly half of the probabilities for S2 and S3 for all of the thresholds explored. However, there remains considerable uncertainty around these results – for example every strategy is dominated by every other strategy in at least some of the PSA samples.

The mean results also indicate that if aripiprazole is used in a three-drug treatment strategy then it may not be optimal if used as a third-line treatment. This strategy is dominated by the strategy of using aripiprazole second-line in 84% of the PSA results, and has a probability of being the most cost-effective strategy of about 0.05 for all of the thresholds explored. The results are less clear about whether or not a specific place in the treatment pathway is indicated for aripiprazole. Use of aripiprazole second-line dominates each of the other strategies in the majority of PSA results. However, for all thresholds up to £95,000, use of aripiprazole third-line has the highest probability of being the most cost-effective treatment strategy.

5.2.11.2 One-way sensitivity analyses (OWSA)

When the model parameters were varied in an OWSA, results were found to be sensitive to the estimated probability of response, in particular the probability of response at week 3 for both aripiprazole and olanzapine. Tables 63 to 65 show the model results when week 3 response is reduced by 30% for aripiprazole and increased by 30% for olanzapine (both separately and collectively).

Table 63: Sensitivity analysis reducing the week 3 probability of response by 30% for aripiprazole

Sequence	Total costs (£)	Total QALYs	Incremental costs (£)	Incremental QALYs	ICER
S1 (RIS, QUE, OLA)	£75,066	2.51637			
S4 (RIS, QUE, ARI)	£75,187	2.51050	£121	-0.00587	Dominated by S1
S2 (RIS, ARI, QUE)	£75,292	2.51083	£227	-0.00554	Dominated by S1
S3 (ARI, RIS, QUE)	£76,760	2.50840	£1,694	-0.00796	Dominated by S1

Table 64: Sensitivity analysis increasing the week 3 probability of response by 30% for olanzapine

Sequence	Total costs (£)	Total QALYs	Incremental costs (£)	Incremental QALYs	ICER
S2 (RIS, ARI, QUE)	£74,133	2.52466			
S3 (ARI, RIS, QUE)	£74,379	2.52348	£246	-0.00118	Dominated by S2
S1 (RIS, QUE, OLA)	£74,763	2.52769	£630	0.00303	£208,149 (Compared with S2)
S4 (RIS, QUE, ARI)	£74,888	2.52297	£124	-0.00472	Dominated by S1

Table 65: Sensitivity analysis reducing the week 3 probability of response by 30% for aripiprazole whilst increasing it by 30% for olanzapine

Sequence	Total costs (£)	Total QALYs	Incremental costs (£)	Incremental QALYs	ICER
S1 (RIS, QUE, OLA)	£74,763	2.52769			
S4 (RIS, QUE, ARI)	£75,187	2.51050	£424	-0.01719	Dominated by S1
S2 (RIS, ARI, QUE)	£75,292	2.51083	£529	-0.01686	Dominated by S1
S3 (ARI, RIS, QUE)	£76,760	2.50840	£1,997	-0.01929	Dominated by S1

Tables 63 and 65 show that a reduction in the effectiveness of aripiprazole by 30% results in all of the treatment strategies containing aripiprazole (S2, S3, S4) becoming dominated by the strategy which excludes aripiprazole (S1). If the effectiveness of olanzapine is increased by 30% (whilst that for aripiprazole is unchanged) then use of aripiprazole third line (S4) is dominated by not using aripiprazole (S1), whilst use of aripiprazole first-line (S3) is dominated by use of aripiprazole second-line (S2).

Within their original submission the manufacturers considered five scenario analyses, displaying the cost-effectiveness results relative to their base case of S1 (treatment without aripiprazole). Four of the results are redisplayed here in Tables 66 to 69, but re-presented as incremental analyses. The scenario analysis considering alternative utility values for weight gain is not considered here as it had a negligible impact on the results.

Table 66: Scenario analyses using 10mg dose for aripiprazole (based on Table B81; MS page 234)

Sequence	Total costs (£)	Total QALYs	Incremental costs (£)	Incremental QALYs	ICER (and comparator)
S2 (RIS, ARI, QUE)	£74,815	2.51991			
S1 (RIS, QUE, OLA)	£75,015	2.51658	£200	-0.00333	Dominated by S2
S4 (RIS, QUE, ARI)	£75,125	2.51865	£310	-0.00126	Dominated by S2
S3 (ARI, RIS, QUE)	£75,741	2.51858	£926	-0.00133	Dominated by S2

Table 67: Scenario analyses using all the identified studies in the network meta-analysis (based on Table B82; MS page 235)

Sequence	Total costs (£)	Total QALYs	Incremental costs (£)	Incremental QALYs	ICER (and comparator)
S2 (RIS, ARI, QUE)	£72,178	2.53391			
S1 (RIS, QUE, OLA)	£72,352	2.53187	£174	-0.00204	Dominated by S2
S4 (RIS, QUE, ARI)	£72,441	2.53284	£263	-0.00107	Dominated by S2
S3 (ARI, RIS, QUE)	£73,164	2.53173	£986	-0.00218	Dominated by S2

Table 68: Scenario analyses using alternative symptom utility values (based on Table B84; MS page 237)

Sequence	Total costs (£)	Total QALYs	Incremental costs (£)	Incremental QALYs	ICER (and comparator)
S2 (RIS, ARI, QUE)	£74,133	2.02172			
S3 (ARI, RIS, QUE)	£74,379	2.0216	£246	-0.00012	Dominated by S2
S4 (RIS, QUE, ARI)	£74,888	2.01838	£755	-0.00334	Dominated by S2
S1 (RIS, QUE, OLA)	£75,066	2.01152	£933	-0.01020	Dominated by S2

Table 69: Scenario analyses changing the order of S1 (based on Table B87; MS page 238)

Sequence	Total costs (£)	Total QALYs	Incremental costs (£)	Incremental QALYs	ICER (and comparator)
S2 (RIS, ARI, OLA)	£73,856	2.52778			
S3 (ARI, RIS, OLA)	£74,102	2.5266	£246	-0.00118	Dominated by S2
S4 (RIS, OLA, ARI)	£74,214	2.5263	£358	-0.00148	Dominated by S2
S1 (RIS, OLA, QUE)	£74,687	2.51672	£831	-0.01106	Dominated by S2

The results of the scenario analyses are consistent in showing that, for a wide variety of model assumptions, use of aripiprazole second-line (after risperidone) dominates the other strategies considered, including not using aripiprazole at all. However, as has previously been shown, there is much uncertainty around this conclusion. In addition, the incremental costs and incremental QALYs are nearly always small, with incremental costs of each strategy (relative to S2) often being less than 2%, and incremental QALYs often being less than 0.05% of the total values for S2.

In response to clarification questions from the ERG, the manufacturers considered additional scenario analyses: varying the starting age of patients, altering treatment duration (acute phase and euthymic: treated phase), increasing relapse rates, reducing treatment efficacy with each additional line of treatment and modelling drug-related adverse event costs. These are discussed in turn.

Varying the starting age of patients

As discussed in Section 5.2.4 of this report, the starting age used in the manufacturers' model was somewhat arbitrary. The ERG requested the manufacturers to explore the impact of using starting ages of 13 years, 17 years and 13.4 years (the mean starting age of patients in the pivotal trial for aripiprazole). Since within the manufacturers' model the starting age has to be an integer, only the first two analyses were performed. Results are reported in Tables 70 and 71. As expected, increasing (or decreasing) the model's time-horizon resulted in increased (or decreased) accumulated costs and QALYs for all the strategies. These results emphasise the fact that the main driver of cost-effectiveness relates to the acute phase of inpatient care.

Table 70: Cost-effectiveness results with a starting age of 13 years (5 year time-horizon)

Sequence	Total costs (£)	Total QALYs	Incremental costs (£)	Incremental QALYs	ICER (and comparator)
S2 (RIS, ARI, QUE)	£97,742	4.06605			
S3 (ARI, RIS, QUE)	£97,975	4.06493	£233	-0.00112	Dominated by S2
S4 (RIS, QUE, ARI)	£98,537	4.06426	£795	-0.00179	Dominated by S2
S1 (RIS, QUE, OLA)	£98,793	4.05384	£1,051	-0.01221	Dominated by S2

Table 71: Cost-effectiveness results with a starting age of 17 years (1 year time-horizon)

Sequence	Total costs (£)	Total QALYs	Incremental costs (£)	Incremental QALYs	ICER (and comparator)
S2 (RIS, ARI, QUE)	£48,388	0.86713			
S3 (ARI, RIS, QUE)	£48,636	0.86594	£248	-0.00119	Dominated by S2
S4 (RIS, QUE, ARI)	£49,027	0.86569	£639	-0.00144	Dominated by S2
S1 (RIS, QUE, OLA)	£49,090	0.86316	£702	-0.00397	Dominated by S2

Altering treatment duration

Based on feedback from their clinical advisors (see Section 5.2.4 of this report), the ERG requested the following additional analyses from the manufacturers:

- An extension of the acute phase of the model beyond three weeks.
- An extension of the euthymic: treated phase of the model to six months.
- An extension of the euthymic: treated phase of the model to twelve months.

In response, the manufacturers extended the acute phase from three weeks to four weeks (question B18, pages 36-37 of clarification response). This had minimal impact on the cost-effectiveness results, and so the results are not re-produced here. The effect of extending the euthymic: treated phase of the model is re-produced as an incremental analysis in Table 72 and Table 73 (based on question B19, pages 37-38 of clarification response).

Table 72: Cost-effectiveness results with an average (mean) of six months euthymic treatment

Sequence	Total costs (£)	Total QALYs	Incremental costs (£)	Incremental QALYs	ICER (and comparator)
S2 (RIS, ARI, QUE)	£68,640	2.53294			
S3 (ARI, RIS, QUE)	£69,203	2.53183	£563	-0.0011	Dominated by S2
S4 (RIS, QUE, ARI)	£69,354	2.52925	£714	-0.0037	Dominated by S2
S1 (RIS, QUE, OLA)	£69,505	2.51989	£865	-0.0130	Dominated by S2

Table 73: Cost-effectiveness results with an average (mean) of twelve months euthymic treatment

Sequence	Total costs (£)	Total QALYs	Incremental costs (£)	Incremental QALYs	ICER (and comparator)
S2 (RIS, ARI, QUE)	£63,994	2.53897			
S3 (ARI, RIS, QUE)	£64,657	2.53348	£663	-0.0055	Dominated by S2
S4 (RIS, QUE, ARI)	£64,773	2.52225	£779	-0.0167	Dominated by S2
S1 (RIS, QUE, OLA)	£64,892	2.53797	£898	-0.0010	Dominated by S2

The results from these additional analyses show that the strategy of using aripiprazole as a second-line treatment dominates each of the other treatment strategies considered by the manufacturers. The ERG notes that increasing the average time spent in the euthymic treated health state reduces the total costs accrued for each treatment strategy.

A further analysis was conducted by the manufacturers, in which the acute phase was extended to four weeks, euthymic treatment was maintained for 12 months, and treatment effectiveness was reduced for second and third-line treatments. The results of this analysis are reported in Table 74.

Increasing relapse rates

Based on feedback from their clinical advisors (see Section 5.2.4 of this report), the ERG requested that the manufacturers explore the effect on their cost-effectiveness results of increasing relapse rates. In response the manufacturers tested rates of 10%, 15% and 20% (over the entire duration of treatment). The results (question B22, pages 42/43 of clarification response) showed that increasing relapse rates resulted in higher accumulated costs and lower

accumulated QALYs for all the strategies, but the incremental analyses remained the same. Only the results for a relapse rate of 20% are presented here, in Table 74.

Table 74: Cost-effectiveness results assuming a total relapse rate of 20% (instead of 5%) whilst euthymic and treated

Sequence	Total costs (£)	Total QALYs	Incremental costs (£)	Incremental QALYs	ICER (and comparator)
S2 (RIS, ARI, QUE)	£104,618	2.47165			
S3 (ARI, RIS, QUE)	£104,766	2.47088	£148	-0.0008	Dominated by S2
S4 (RIS, QUE, ARI)	£105,241	2.47038	£623	-0.0013	Dominated by S2
S1 (RIS, QUE, OLA)	£106,176	2.4629	£1,558	-0.0088	Dominated by S2

Reducing treatment efficacy with each additional line of treatment

The ERG requested that the manufacturers explore the possibility that treatment efficacy reduces when the antipsychotic is not used as first line treatment (see Section 5.2.6 of this report). In response the manufacturers conducted the following analyses:

- A reduction in efficacy of 5% between lines 1 and 2, and 10% between lines 2 and 3.
- A reduction in efficacy of 10% between lines 1 and 2, and 15% between lines 2 and 3.
- A reduction in efficacy of 15% between lines 1 and 2, and 20% between lines 2 and 3.
- A reduction in efficacy of 50% between lines 1 and 2, and 75% between lines 2 and 3.

The results (question B16, pages 35-36 of clarification response) show that the manufacturers' original cost-effectiveness results are not substantially altered – even with reductions of 50% and 75%. Only results for reductions of 10% and 15% are reproduced in Table 75.

Table 75: Cost-effectiveness results assuming a reduction in efficacy of 10% between lines 1 and 2, and 15% between lines 2 and 3

Sequence	Total costs (£)	Total QALYs	Incremental costs (£)	Incremental QALYs	ICER (and comparator)
S2 (RIS, ARI, QUE)	£74,881	2.51736			
S3 (ARI, RIS, QUE)	£75,142	2.51602	£261	-0.0013	Dominated by S2
S4 (RIS, QUE, ARI)	£75,608	2.51523	£727	-0.0021	Dominated by S2
S1 (RIS, QUE, OLA)	£75,777	2.50872	£896	-0.0086	Dominated by S2

The manufacturers also conducted a sensitivity analysis in which the analysis of Table 75 was combined with the additional assumptions that the acute phase was extended to four weeks and euthymic treatment was maintained for 12 months. The results are reproduced in Table 76 (from question B20, page 38 of clarification response).

Table 76: Cost-effectiveness results assuming a reduction in efficacy of 10% between lines 1 and 2, and 15% between lines 2 and 3, four weeks of acute treatment and 12 months of euthymic treatment

Sequence	Total costs (£)	Total QALYs	Incremental costs (£)	Incremental QALYs	ICER (and comparator)
S2 (RIS, ARI, QUE)	£63,994	2.53897			
S3 (ARI, RIS, QUE)	£64,657	2.53348	£663	-0.0055	Dominated by S2
S4 (RIS, QUE, ARI)	£64,773	2.52225	£779	-0.0167	Dominated by S2
S1 (RIS, QUE, OLA)	£64,892	2.53797	£898	-0.001	Dominated by S2

Modelling drug-related adverse event costs

In their original submission the manufacturers assumed that drug-related adverse events experienced by inpatients would be captured by their use of NHS Reference Costs. However, because the incidence of drug-related adverse events varies by drug, the ERG requested that the manufacturers explicitly model this (see Section 5.2.6 of this report). Due to uncertainty in the resource use for treating EPS and somnolence, the manufacturers considered two different scenarios:

- Somnolence and EPS both requiring one additional hour of consultant time per week.
- Somnolence and EPS both requiring three additional hours of consultant time per week.

The results of these analyses are reproduced in Tables 77 and 78 (from question B21, pages 39-41 of clarification response). They show that use of aripiprazole, second-line, continues to dominate all of the other treatment sequences considered. The ERG's clinical advisors stated that the time required spent treating EPS and somnolence adverse events would not be excessive and that the one additional hour of consultant time per week assumed in Table 77 is likely to over-estimate the time required.

Table 77: Cost-effectiveness results assuming that somnolence and EPS both require one additional hour of consultant time per week

Sequence	Total costs (£)	Total QALYs	Incremental costs (£)	Incremental QALYs	ICER (and comparator)
S2 (RIS, ARI, QUE)	£75,760	2.52466			
S3 (ARI, RIS, QUE)	£76,198	2.52348	£438	-0.0012	Dominated by S2
S1 (RIS, QUE, OLA)	£76,425	2.51637	£665	-0.0083	Dominated by S2
S4 (RIS, QUE, ARI)	£76,453	2.52297	£693	-0.0017	Dominated by S2

Table 78: Cost-effectiveness results assuming that somnolence and EPS both require three additional hours of consultant time per week

Sequence	Total costs (£)	Total QALYs	Incremental costs (£)	Incremental QALYs	ICER (and comparator)
S2 (RIS, ARI, QUE)	£79,014	2.52466			
S3 (ARI, RIS, QUE)	£79,143	2.51637	£129	-0.0083	Dominated by S2
S1 (RIS, QUE, OLA)	£79,584	2.52297	£570	-0.0017	Dominated by S2
S4 (RIS, QUE, ARI)	£79,834	2.52348	£820	-0.0012	Dominated by S2

5.3 Additional work undertaken by the ERG

The ERG noted that because aripiprazole only has positive CHMP opinion for 12 weeks of treatment it would not be appropriate for the manufacturers to promote treatment with aripiprazole beyond 12 weeks. Hence the ERG carried out an initial analysis in which none of the antipsychotics were used for more than 12 weeks in any treatment line recognising that the comparator interventions are used off-label. The results are reproduced in Table 79 and show that limiting treatment to 12 weeks has little impact on the manufacturers' original cost-effectiveness results.

Table 79: Cost-effectiveness results limiting antipsychotic treatment to a maximum of 12 weeks per treatment line

Sequence	Total costs (£)	Total QALYs	Incremental costs (£)	Incremental QALYs	ICER (and comparator)
S2 (RIS, ARI, QUE)	£73,673	2.52609			
S3 (ARI, RIS, QUE)	£73,897	2.5249	£224	-0.0012	Dominated by S2
S4 (RIS, QUE, ARI)	£74,428	2.52456	£755	-0.0015	Dominated by S2
S1 (RIS, QUE, OLA)	£74,673	2.51832	£999	-0.0078	Dominated by S2

5.3.1 Exploratory analysis of the impact of personalised-medicine

As has been previously noted, clinical advisors to the ERG and clinical advisors to the manufacturers both pointed out that the specific treatment adopted will be tailored to the individual based on a combination of factors including: severity of symptoms; side-effect profile; comorbidities; and the likelihood of adherence. This suggests that certain sub-groups may benefit more than others from the use of specific antipsychotic treatment sequences. Examples described to the ERG include the use of olanzapine for highly irritable/agitated patients and the use of quetiapine for depressed patients (Section 3 of this report). There are limited data available to model treatment effects within sub-groups, hence the ERG conducted an exploratory scenario analysis to examine the possible implications of this personalised medicine.

For this analysis, the ERG adopted the results presented in Table 76 of this report as their base case. For ease of readability these results are reproduced in Table 80 which presents the results of amending the manufacturers' original analysis by assuming:

- A reduction in efficacy of 10% between lines 1 and 2, and 15% between lines 2 and 3
- Four weeks of acute treatment
- Twelve months of euthymic treatment.

Table 80: The base case used in the exploratory analysis of personalised medicine

Sequence	Total costs (£)	Total QALYs	Incremental costs (£)	Incremental QALYs	ICER (and comparator)
S2 (RIS, ARI, QUE)	£63,994	2.53897			
S4 (RIS, QUE, ARI)	£64,657	2.53348	£663	-0.0055	Dominated by S2
S1 (RIS, QUE, OLA)	£64,773	2.52225	£779	-0.0167	Dominated by S2
S3 (ARI, RIS, QUE)	£64,892	2.53797	£898	-0.001	Dominated by S2

The potential impact of personalised-medicine was modelled either by applying an additional QALY gain or by offsetting the total costs for S1, S3 and S4 (no change was applied to S2 as this would not change the base case results). The impact on the resulting ICER of adding different QALY gains and different costs offsets was explored and is presented in Figures 12 to 14. These graphs all use the same axis scale. The QALY gains and cost offsets required for each of the strategies (S1, S3 and S4) to produce an ICER (relative to S2) equal to £30,000 and to £20,000 are displayed in Table 81.

Figure 12: Effect on the ICER (vs. S2) of applying additional QALY gains to S4

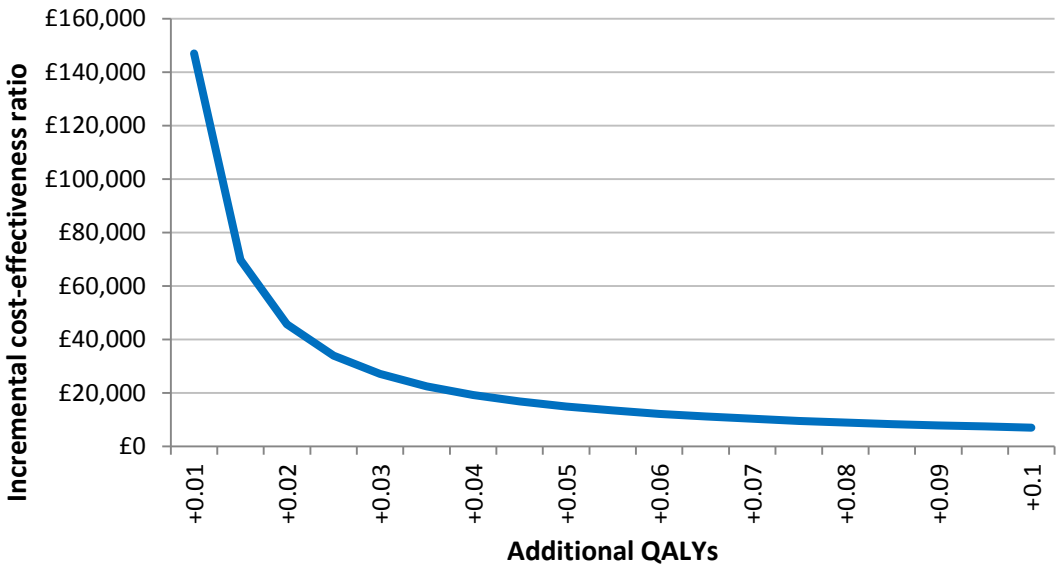


Figure 13: Effect on the ICER (vs. S2) of applying additional QALY gains to S1

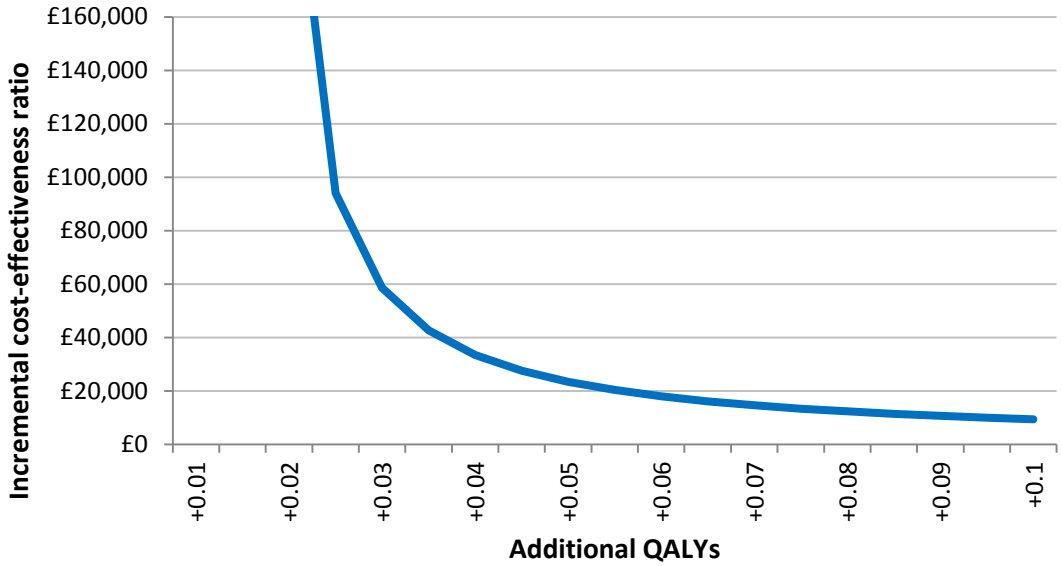


Figure 14: Effect on the ICER (vs. S2) of applying additional QALY gains to S3

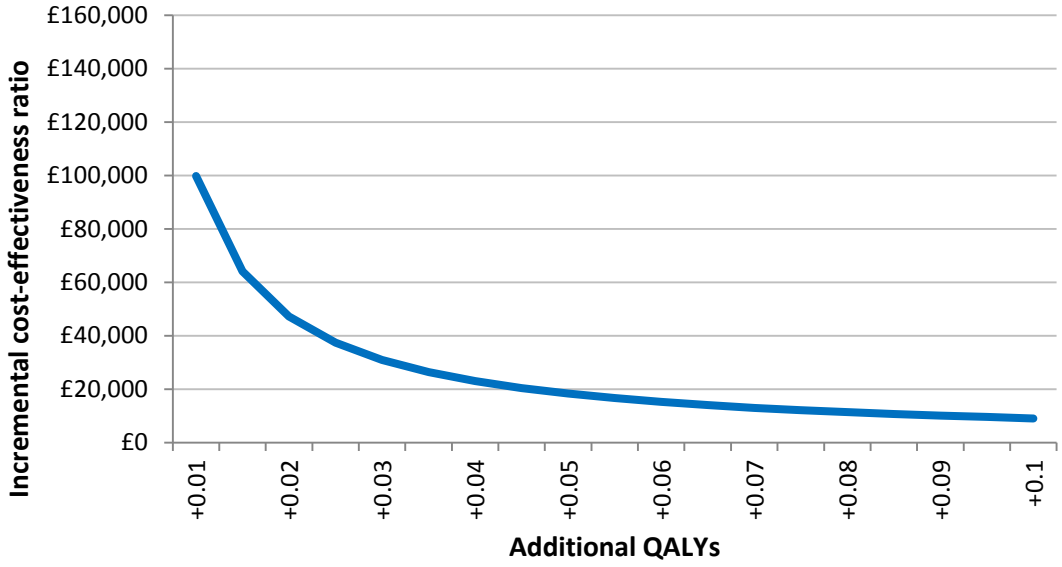


Figure 15: Effect on the ICER (vs. S2) of applying additional cost reductions to S4

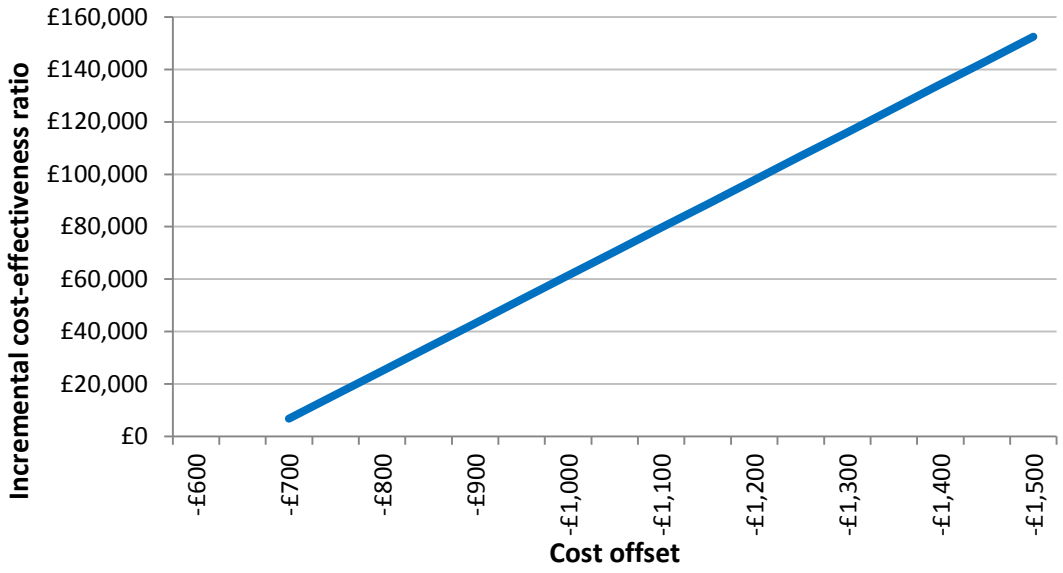


Figure 16: Effect on the ICER (vs. S2) of applying additional cost reductions to S1

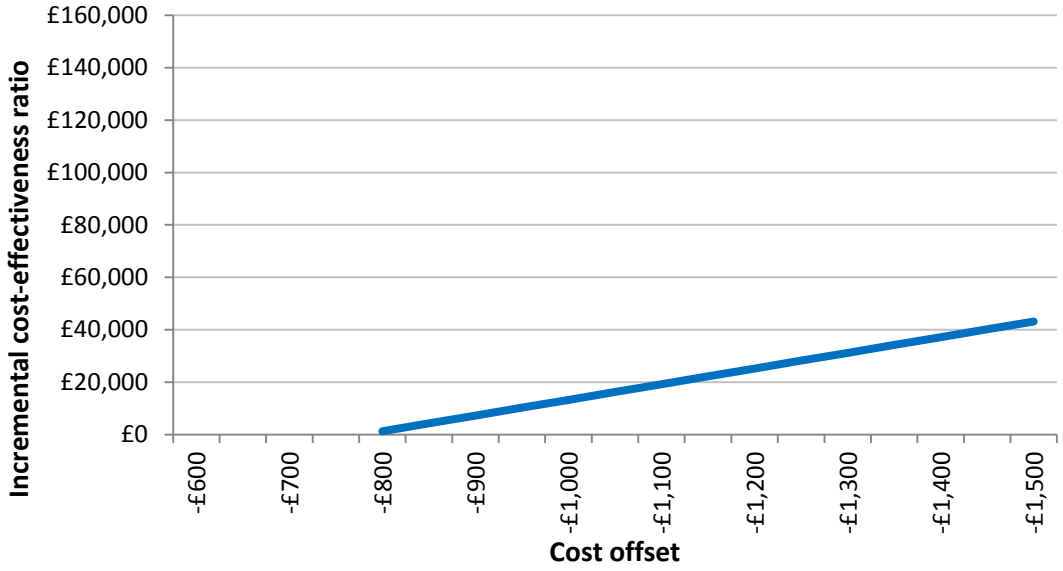


Figure 17: Effect on the ICER (vs. S2) of applying additional cost reductions to S3

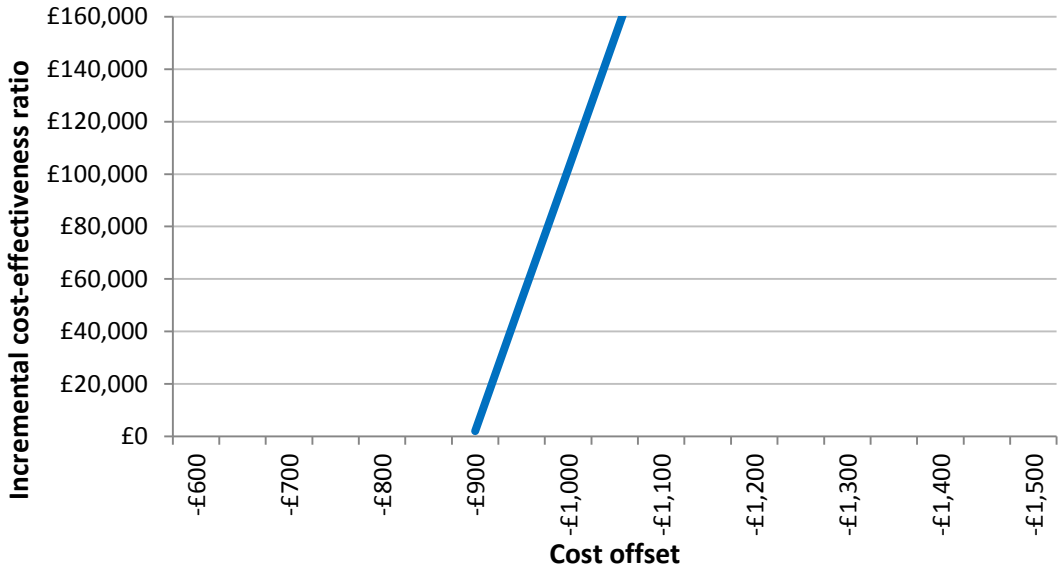


Table 81: QALY gains and cost offsets required to produce ICER equal to £30,000 or £20,000 per QALY gained (relative to S2)

Strategy	Additional QALYs required for:		Cost offset required for	
	ICER = £30,000	ICER = £20,000	ICER = £30,000	ICER = £20,000
S4 (RIS, QUE, ARI)	0.02759	0.03864	- £828	- £773
S1 (RIS, QUE, OLA)	0.04269	0.05567	- £1,281	- £1,113
S3 (ARI, RIS, QUE)	0.03093	0.04590	- £928	- £918

These results indicate that the increased QALYs or cost-offsets required by strategies 1, 3 and 4 to have an ICER below £20,000 or £30,000 per QALY gained are not excessive. It is commented that the assumed cost of a weeks' stay as an inpatient is estimated to be greater than £4,200.

The calculations presented in Table 69 do not take into consideration simultaneous cost offsets and QALY gains associated with more effective personalised treatments.

5.3.2 The ERG's estimation of a plausible ICER

The ERG is generally satisfied with the base case analysis presented by the manufacturers. However, the ERG considers that the following changes should be made to generate a more plausible ICER:

i. Include a half-cycle correction

ii. Amend the discounting calculations (see Section 5.2.4)

iii. Amend the mortality calculations (see Section 5.2.6)

iv. A reduction in efficacy of 10% between lines 1 and 2, and 15% between lines 2 and 3

v. Using a random-effects model to conduct the network meta-analysis (see Section 4.4).

vi. Restrict the network meta-analysis results used in the PSA such that the week 3 probability of discontinuing or responding does not exceed 100% (see Section 5.2.11.1)

In addition, the ERG notes that there are two possible ways that treatment with aripiprazole could be modelled:

A) The model could reflect the licenced duration of 12 weeks for aripiprazole, giving a 'licenced duration' model.

B) The model could reflect real-world prescribing of aripiprazole. This would give a 'real-world' model.

Both model types were explored for the ERG's plausible ICER. For the real-world model the following treatment lengths were used:

- Four weeks of acute treatment
- Twelve months of euthymic treatment.

The results for the licenced duration model are presented in Table 82 (deterministic results) and Table 83 (probabilistic results), whilst the results from the real-world model are presented in Table 84 (deterministic results) and Table 85 (probabilistic results). The results show close agreement between the deterministic and probabilistic results for both of the models.

Table 82: Deterministic cost-effectiveness results based on the ERG's amendments: licenced duration model

Sequence	Total costs (£)	Total QALYs	Incremental costs (£)	Incremental QALYs	ICER (and comparator)
S2 (RIS, ARI, QUE)	£70,647	2.46858			
S3 (ARI, RIS, QUE)	£70,821	2.46756	£174	-0.00101	Dominated by S2
S4 (RIS, QUE, ARI)	£71,393	2.46706	£747	-0.00152	Dominated by S2
S1 (RIS, QUE, OLA)	£72,411	2.45340	£1,764	-0.01518	Dominated by S2

Table 83: Probabilistic cost-effectiveness results based on the ERG's amendments: licenced duration model

Sequence	Total costs (£)	Total QALYs	Incremental costs (£)	Incremental QALYs	ICER (and comparator)
S2 (RIS, ARI, QUE)	£70,707	2.47064			
S3 (ARI, RIS, QUE)	£70,881	2.46972	£174	-0.00092	Dominated by S2
S4 (RIS, QUE, ARI)	£71,454	2.46883	£747	-0.00181	Dominated by S2
S1 (RIS, QUE, OLA)	£72,157	2.45798	£1,450	-0.01267	Dominated by S2

Table 84: Deterministic cost-effectiveness results based on the ERG's amendments: real-world model

Sequence	Total costs (£)	Total QALYs	Incremental costs (£)	Incremental QALYs	ICER (and comparator)
S2 (RIS, ARI, QUE)	£62,257	2.48639			
S4 (RIS, QUE, ARI)	£63,065	2.48062	£808	-0.00576	Dominated by S2
S1 (RIS, QUE, OLA)	£63,293	2.47160	£1,035	-0.01478	Dominated by S2
S3 (ARI, RIS, QUE)	£63,437	2.48511	£1,180	-0.00127	Dominated by S2

Table 85: Probabilistic cost-effectiveness results based on the ERG's amendments: real-world model

Sequence	Total costs (£)	Total QALYs	Incremental costs (£)	Incremental QALYs	ICER (and comparator)
S2 (RIS, ARI, QUE)	£62,138	2.42890			
S4 (RIS, QUE, ARI)	£62,880	2.42301	£742	-0.00589	Dominated by S2
S1 (RIS, QUE, OLA)	£63,051	2.41584	£912	-0.01306	Dominated by S2
S3 (ARI, RIS, QUE)	£63,384	2.42797	£1,245	-0.00093	Dominated by S2

Compared with the manufacturers' deterministic base case results (reproduced in Table 57 of this report), the licenced duration model results show a decrease in both total costs and total QALYs. This decrease is consistent across the four strategies considered, with the exception of S1 which has less favourable outcomes for both costs and QALYs relative to S2.

Compared with the manufacturers' deterministic base case results (reproduced in Table 57 of this report), the real-world model results show a decrease in both total costs and total QALYs. This decrease is inconsistent across the four strategies considered, with strategy S2 becoming more favourable relative to each of the other strategies.

5.4 Conclusions

The manufacturers' report was well written and their model was transparent and well structured. The manufacturers provided amended models in response to all of the ERG's clarification questions, with the exception of including a model which considered four treatment lines of antipsychotics. This was not provided due to a lack of time during the clarification process, although the manufacturers indicated that they would be able to provide this if required.

Whilst there are a large number of uncertainties relating to the manufacturers' economic evaluation, many of these do not have an appreciable impact on the cost-effectiveness results. Based on deterministic results, the base case conclusion that aripiprazole second-line (S2) dominates each of the other treatment strategies is only changed in the following situations:

- **The week 3 probability of YMRS response is reduced by 30% for aripiprazole**. The strategy excluding aripiprazole (S1) then dominates all of the strategies that include aripiprazole – see Table 60 of this report.
  This occurs in 2 (0.2%) of the manufacturers' PSA runs and in 53 (5.3%) of the ERG's PSA runs.
- **The week 3 probability of YMRS response is increased by 30% for olanzapine**. S1 has an ICER of £208,149 compared with S2. Use of aripiprazole first line (S3) or third line (S4) is dominated by S2 and S1 respectively – see Table 65 of this report.
  This occurs in 39 (3.9%) of the manufacturers' PSA runs and in 85 (8.5%) of the ERG's PSA runs.

However, the PSA results indicate that there is considerable uncertainty over the conclusion that S2 dominates all of the other treatment strategies. Using the manufacturers' base case analysis, the probability of S2 not being the most-cost effective strategy exceeds 60% for willingness-to-pay thresholds between £0 per QALY and £100,000 per QALY. In addition, 95% confidence intervals about the deterministic results include the possibility that S2 is dominated by each of the other strategies.

The possibility that any of the four treatment strategies considered may represent cost-effective options is further emphasised by the ERG's additional exploratory work into personalised medicine (Section 5.3.1). This shows that changes of between 1% and 2% in the base case results (costs or QALYs) for S1, S3 and S4 can bring the ICERs (relative to S2) below £30,000.

A remaining, unresolved uncertainty is the impact on the cost-effectiveness results of using four treatment lines of antipsychotics. The ERG, based on clinical advice, note that the use of four treatment lines within the manufacturers' model would make it more realistic.

The manufacturers and the ERG disagreed over the form of model that should be used within the network meta-analysis. The manufacturers favoured a fixed-effects model, claiming that there was not enough evidence to conduct a random-effects model. The ERG considers that it is more important to explore uncertainty when there is a lack of evidence, and so favour a random-effects model. It is noted that the point-estimates (and so estimates of deterministic cost-effectiveness) remain largely unchanged between the two models, but the use of a random-effects model introduces additional uncertainty into the results of the economic evaluation.

In the MS the only included adverse-event costs were for weight gain. The ERG notes that this is likely to create a slight bias in favour of aripiprazole as it has the highest incidence of both EPS and somnolence, along with the lowest incidence of weight gain. However, in response to clarification (question B21, pages 39-41) the manufacturers demonstrated that the base case cost-effectiveness results are not substantially altered when costs are included for all of the adverse-events. In addition, the ERG notes that, due to a lack of evidence, the manufacturers did not model the effects of prolactin increase, and they made conservative assumptions about the incidence of EPS and somnolence for olanzapine. Both of these will create a small bias disfavouring aripiprazole.

6 IMPACT ON THE ICER OF ADDITIONAL CLINICAL AND ECONOMIC ANALYSES UNDERTAKEN BY THE ERG

In order to provide an indication of the key drivers to the change in the ICER, the amendments made in the ERG's two models were made independently to the manufacturers' deterministic base case. Amending the model so that the combined probability of response of discontinuation does not exceed 100% was not undertaken as the summation of response and discontinuation was below 100% in the deterministic scenario. For all of the amendments, S2 dominates each of the other strategies, so only incremental costs and QALYs are displayed (both relative to S2). Results are displayed in Tables 86 and 87.

Table 86: Changes in incremental costs based on the ERG amendments

Incremental Costs (relative to S2)	S1	S3	S4
Manufacturers' base case	£932	£246	£754
Including a half-cycle correction	£932	£197	£754
Using the ERG random effects (instead of fixed effects) network meta-analyses	£950	£320	£756
Amending discounting calculations to be weekly instead of yearly	£921	£245	£746
Amending mortality calculations so that mortality rate cannot exceed 100%	£932	£246	£754
Reduction in efficacy of 10% between lines 1 and 2, and 15% between lines 2 and 3 (base case: no reductions)	£896	£261	£727
Twelve weeks maximum treatment (base case: twelve weeks average treatment)	£999	£224	£755
Four weeks acute treatment instead of three	£1,276	£444	£963
Twelve months of euthymic treatment instead of an average of 10 to 12 weeks.	£779	£898	£663
ERG base case (licensed prescribing)	£1,764	£174	£747
ERG base case (real-world prescribing)	£1,035	£1,180	£808

Table 87: Changes in incremental QALYs based on the ERG amendments

Incremental QALYs (relative to S2)	S1	S3	S4
Manufacturers' base case	-0.0083	-0.0012	-0.0017
Including a half-cycle correction	-0.0083	-0.0010	-0.0017
Using the ERG random effects (instead of fixed effects) network meta-analyses	-0.0084	-0.0012	-0.0018
Amending discounting calculations to be weekly instead of yearly	-0.0082	-0.0012	-0.0017
Amending mortality calculations so that mortality rate cannot exceed 100%	-0.0083	-0.0012	-0.0017
Reduction in efficacy of 10% between lines 1 and 2, and 15% between lines 2 and 3 (base case: no reductions)	-0.0086	-0.0013	-0.0021
Twelve weeks maximum treatment (base case: twelve weeks average treatment)	-0.0078	-0.0012	-0.0015
Four weeks acute treatment instead of three	-0.0077	-0.0018	-0.0021
Twelve months of euthymic treatment instead of an average of 10 to 12 weeks.	-0.0167	-0.0010	-0.0055
ERG base case (licensed prescribing)	-0.0152	-0.0010	-0.0015
ERG base case (real-world prescribing)	-0.0148	-0.0013	-0.0058

The results show that of the different amendments, only the two that extend treatment duration (to four weeks acute treatment or twelve months of euthymic treatment) substantially alter the incremental results, in general making S2 slightly more favourable than the other treatment strategies.

7 END OF LIFE

Aripiprazole does not meet the end of life criteria published by NICE. Although the intervention is anticipated to be indicated for a small patient population, it is not indicated for patients with a short life expectancy and there is no evidence that the intervention offers an extension to life.

8 CONCLUSIONS

On the basis of the clinical evidence provided in the MS, aripiprazole has a similar efficacy profile, in terms of YMRS reduction, as the comparator antipsychotics: olanzapine; risperidone; and quetiapine. There is no conclusive evidence that aripiprazole has a worse side effect profile than olanzapine, risperidone and quetiapine (although the point estimate for rates of somnolence is higher). There exists evidence that aripiprazole may have a reduced incidence of clinically significant weight gain and clinically significant increase in prolactin levels compared with the comparator antipsychotics.

The ERG considers the US paediatric bipolar I population included in trial NCT00110461 to be discrepant to UK population according to the low mean age and high prevalence of comorbid ADHD. Additionally, the severity of the patients included in the trial population in the MS is unlikely to reflect clinical practice in the UK. This is due to the inclusion criteria employed in trial NCT00110461[21] stipulating that suicidal patients were excluded from participating in the study. Furthermore, the manufacturers were unable to provide the ERG with data on the number of trial patients who were inpatients (as would be the case in UK clinical practice) which also suggests that the population in the MS may not reflect the UK paediatric bipolar I population.

As the NCT00110461 trial[21] duration was 30 weeks, the duration of maintenance of effect of only 12 weeks of aripiprazole treatment is unknown. No recurrence data were provided by the manufacturers to indicate how long patients in the included trial remain stable following discontinuation of antipsychotic treatment. The focus of the MS was treatment of the acute phase. The use of aripiprazole as maintenance therapy, as may be used in clinical practice, is outside the CHMP's recommended duration of treatment. However, the 30 week data indicate that the safety profile of aripiprazole during the extension phase was acceptable.

It is noted that not all the information requested by the ERG were made available. It is unclear whether, if these data were known, this would have an impact on the clinical interpretation. Data on adherence was collected but not provided in the MS. The categories for which incomplete information was provided included: comorbid ADHD; the numbers in age subgroups; rapid cyclers; mixed/ manic episode. No data were provided on the numbers in receipt of psychotherapy; and the number of patients in community versus inpatient care.

Within the MS it was stated that the use of aripiprazole at any point in a treatment sequence is a cost-effective alternative to not using aripiprazole. This conclusion was not based on fully

incremental analyses (instead each treatment strategy including aripiprazole was compared with the treatment strategy excluding aripiprazole). The ERG performed incremental analyses and found that use of aripiprazole second-line dominated its use at any other point in the treatment sequence (including not being used). However, the ERG also performed additional work looking at the uncertainty in this conclusion and the potential impact of personalised medicine, as the clinical advisors to the ERG stated that their choice of first-line treatment varied depending on the patient. These results suggested that it is possible that the optimal treatment sequence could depend on patient characteristics. This conclusion is not in disagreement with the manufacturers' conclusion.

In addition, the ERG considered two different treatment durations for aripiprazole (and the other antipsychotics) relating to the CHMP guidance and the current use of antipsychotics as detailed to the ERG by its clinical advisors. The conclusions from these two durations did not differ.

It is noted that the MS did not consider four lines of antipsychotic treatments which would allow a patient to be prescribed each of aripiprazole, olanzapine, risperidone, and quetiapine as necessary. The ERG's clinical advisors believed this was more likely to be the case than only using three antipsychotics. It is not believed that the inclusion of a fourth treatment line would alter the conclusions.

In summary, both the conclusions of the manufacturers and the ERG are that the addition of aripiprazole (at the expense of olanzapine or quetiapine) is likely to be beneficial. A fully incremental analysis, using average costs and QALYs indicates that a strategy of using aripiprazole second-line following risperidone would be the most cost-effective strategy, although this may alter were an individual patient to have more QALYs and / or lesser costs under a particular strategy.

8.1 Implications for research

The following areas were identified as being worthy of future research:

- Efficacy and safety data in sample representative of UK paediatric patients with bipolar I disorder.
- Changes in the effectiveness of each antipsychotic when used first, second or third line.

- Measurement of preference-based health-related quality of life for patients with paediatric bipolar disorder.

9 Appendix 1 ERG search strategies

Embase search strategy (340 records)

1. exp Bipolar Disorder/
2. ((bipolar or bi polar) adj5 (disorder$ or depress$)).tw.
3. (hypomania$ or mania$ or manic$).tw.
4. (((cyclothymi$ or rapid or ultradian) adj5 cycl$) or RCBD).tw.
5. or/1-4
6. exp clinical trials/ or exp clinical trial/ or exp controlled clinical trials/
7. exp crossover procedure/ or exp cross over studies/ or exp crossover design/
8. exp double blind procedure/ or exp double blind method/ or exp double blind studies/
9. exp single blind procedure/ or exp single blind method/ or exp single blind studies/
10. exp random allocation/ or exp randomization/ or exp random sample/
11. exp randomized controlled trials/ or exp randomized controlled trial/
12. (clinical adj2 trial$).tw.
13. (((single$ or doubl$ or trebl$ or tripl$) adj5 (blind$ or mask$ or dummy)) or (singleblind$ or doubleblind$ or trebleblind$)).tw.
14. (placebo$ or random$).mp.
15. exp epidemiologic study characteristics/
16. animals/ not (animals/ and human$.mp.)
17. animal$/ not (animal$/ and human$/)
18. (or/6-15) not (or/16-17)
19. exp child/ or exp adult children/ or exp adolescent/
20. (child$ or adolescen$ or youth$ or preschool or juvenile or pediatric or paediatric).tw.
21. (young adj3 (person$ or people)).tw.
22. under 18.tw.
23. under eighteen.tw.
24. or/19-23
25. 5 and 18
26. 24 and 25
27. limit 26 to yr="2005 -Current"
28. exp Bipolar Disorder/
29. ((bipolar or bi polar) adj5 (disorder$ or depress$)).tw.
30. (hypomania$ or mania$ or manic$).tw.
31. (((cyclothymi$ or rapid or ultradian) adj5 cycl$) or RCBD).tw.
32. or/28-31
33. exp child/ or exp adult children/ or exp adolescent/
34. (child$ or adolescen$ or youth$ or preschool or juvenile or pediatric or paediatric).tw.
35. (young adj3 (person$ or people)).tw.
36. under 18.tw.
37. under eighteen.tw.
38. or/33-37
39. 32 and 38
40. limit 39 to yr="2005 -Current"
41. ((side or adverse or undesirable) adj2 (event$ or effect$ or reaction$ or outcome$)).ti.
42. (safe or safety).ti.
43. (harm$ or complication$).ti.
44. risk$.ti.
45. (treatment adj emergen$).ti.
46. tolerability.ti.
47. mortality.ti.
48. or/41-47

49. 40 and 48
50. aripiprazole/ae, to [Adverse Drug Reaction, Drug Toxicity]
51. aripiprazole.ti,ab.
52. 50 or 51
53. 40 and 52
54. extrapyramidal symptom/
55. (extrapyramidal symptom$ or EPS).ti,ab.
56. 54 or 55
57. weight gain/
58. weight gain$.ti,ab.
59. 57 or 58
60. somnolence/
61. somnolence.ti,ab.
62. 60 or 61
63. nausea/
64. nause$.ti,ab.
65. 63 or 64
66. 56 or 59 or 62 or 65
67. 40 and 66
68. dopamine receptor stimulating agent/ae, to [Adverse Drug Reaction, Drug Toxicity]
69. 40 and 68
70. 49 or 53 or 67 or 69
71. 70 not 27

Medline search strategy (254 records)

1. Bipolar Disorder/
2. ((bipolar or bi polar) adj5 (disorder$ or depress$)).tw.
3. (hypomania$ or mania$ or manic$).tw.
4. (((cyclothymi$ or rapid or ultradian) adj5 cycl$) or RCBD).tw.
5. or/1-4
6. exp clinical trial/ or exp controlled clinical trials/
7. exp cross over studies/ or exp crossover design/
8. exp double blind method/ or exp double blind studies/
9. exp single blind method/ or exp single blind studies/
10. exp random allocation/ or exp randomization/
11. exp randomized controlled trials/ or exp randomized controlled trial/
12. (clinical adj2 trial$).tw.
13. (((single$ or doubl$ or trebl$ or tripl$) adj5 (blind$ or mask$ or dummy)) or (singleblind$ or doubleblind$ or trebleblind$)).tw.
14. (placebo$ or random$).mp.
15. (clinical trial$ or clinical control trial or random$).pt.
16. exp epidemiologic study characteristics/
17. animals/ not (animals/ and human$.mp.)
18. animal$/ not (animal$/ and human$/)
19. (or/6-15) not (or/17-18)
20. exp child/ or exp adult children/ or exp adolescent/
21. (child$ or adolescen$ or youth$ or preschool or juvenile or pediatric or paediatric).tw.
22. (young adj3 (person$ or people)).tw.
23. under 18.tw.
24. under eighteen.tw.
25. or/20-24
26. 5 and 19

27. 25 and 26
28. Bipolar Disorder/
29. ((bipolar or bi polar) adj5 (disorder$ or depress$)).tw.
30. (hypomania$ or mania$ or manic$).tw.
31. (((cyclothymi$ or rapid or ultradian) adj5 cycl$) or RCBD).tw.
32. or/28-31
33. exp child/ or exp adult children/ or exp adolescent/
34. (child$ or adolescen$ or youth$ or preschool or juvenile or pediatric or paediatric).tw.
35. (young adj3 (person$ or people)).tw.
36. under 18.tw.
37. under eighteen.tw.
38. or/33-37
39. 32 and 38
40. limit 39 to yr="2005 -Current"
41. ((side or adverse or undesirable) adj2 (event$ or effect$ or reaction$ or outcome$)).ti.
42. (safe or safety).ti.
43. (harm$ or complication$).ti.
44. risk$.ti.
45. (treatment adj emergen$).ti.
46. tolerability.ti.
47. mortality.ti.
48. or/41-47
49. 40 and 48
50. (extrapyramidal symptom$ or EPS).ti,ab.
51. Weight Gain/
52. weight gain$.ti,ab.
53. 51 or 52
54. somnolence.ti,ab.
55. Nausea/
56. nause$.ti,ab.
57. 55 or 56
58. 50 or 53 or 54 or 57
59. 40 and 58
60. Dopamine Agonists/ae, to [Adverse Effects, Toxicity]
61. 40 and 60
62. 49 or 59 or 61
63. 62 not 27

10 REFERENCES

1. Tramontina, S., Zeni, C., Ketzer, C., Pheula, G., Narvaez, J., Rohde, L. Aripiprazole in children and adolescents with bipolar disorder comorbid with attention-deficit/hyperactivity disorder: a pilot randomized clinical trial. *J Clin Psychiatry* 2009; 70:756-764.

2. National Institute of Mental Health. Bipolar Disorder in Children and Teens: A Parent's Guide . 2008.

3. National Clinical Practice Guideline. Bipolar disorder. The management of bipolar disorder in adults, children and adolescents, in primary and secondary care. *National Collaborating Centre for Mental Health* 2006; 38.

4. Youngstrom, E.A., Freeman, A.J., Jenkins, M.M. The assessment of children and adolescents with bipolar disorder. *Child Adolesc Psychiatr Clin N Am* 2009; 18(2):353-390.

5. DelBello, M., Kowatch, R., Adler, C., Stanford, K.E., Welge, J.A., Barzman, D.H. et al. A Double-Blind Randomized Pilot Study Comparing Quetiapine and Divalproex for Adolescent Mania. *J Am Acad Child Adolesc Psychiatry* 2006; 45(3):305-313.

6. Freeman, A., Youngstrom, E., Michalek, E., Siegel, R., Meyers, O., Findling, R. Quality of Life in Pediatric Bipolar Disorder. *Pediatrics* 2013; 123:e446-e452.

7. Rademacher, J., DelBello, M.P., Adler, C., Stanford, K., Strakowski, S.M. Health-Related Quality of Life in Adolescents with Bipolar I Disorder. *Journal of Child and Adolescent Psychopharmacology* 2007; 17(1):97-103.

8. Soutullo, C., Chang, K., Diez-Suarez, A., Figueroa-Quintana, A., Escamilla-Canales, I., Rapado Castro, M. et al. Bipolar disorder in children and adolescents: international perspective on epidemiology and phenomenology. *Bipolar Disorders* 2005; 7:497-506.

9. O'Herlihy, A., Worrall, A., Banerjee, S., Jaffa, T., Hill, P., Mears, A. et al. National In-patient Child and Adolescent Psychiatry Study (NICAPS). 2001.

10. Stringaris, A., Santosh, P., Leibenluft, E., Goodman, R. Youth meeting symptom and impairment criteria for maniaGÇÉlike episodes lasting less than four days: an epidemiological enquiry. *Journal of Child Psychology and Psychiatry* 2009; 51(1):31-38.

11. Pini, S., de Queiroz, V., Pagnin, D., Pezawas, L., Angst, J., Cassano, G.B. et al. Prevalence and burden of bipolar disorders in European countries. *European Neuropsychopharmacology* 2005; 15(4):425-434.

12. Merikangas, K.R., Lamers, F. The 'true' prevalence of bipolar II disorder. *Current Opinion in Psychiatry* 2012; 25(1):19-23.

13. Moreno, D.H., Andrade, L.H. Latent class analysis of manic and depressive symptoms in a population-based sample in Sao Paulo, Brazil. *Journal of Affective Disorders* 2010; 123(1-3):208-215.

14. Youngstrom, E.A., Freeman, A.J., McKeown Jenkins, M. The Assessment of Children and Adolescents with Bipolar Disorder. *Child and Adolescent Psychiatric Clinics of North America* 2009; 18(2):353-390.

15. Cahill, C., Hanstock, T., Jairam, R., Hazell, P., Walter, G., Malhi, G.S. Comparison of diagnostic guidelines for juvenile bipolar disorder. *Australasian Psychiatry* 2007; 41(6):479-484.

16. Marchand, WR., Wirth, L., Simon, C. Delayed Diagnosis of Pediatric Bipolar Disorder in a Community Mental Health Setting. *Journal of Psychiatric Practice-«* 2006; 12(2).

17. Royal College of Paediatrics and Child Health. The Use Of Unlicensed Medicines Or Licensed Medicines For Unlicensed Applications In Paediatric Practice: Policy statement produced by the joint RCPCH/NPPG Standing Committee on Medicines. 2000;(accessed Feb. 2013).

18. Cipriani, A., Barbui, C., Salanti, G., Rendell, J., Brown, R., Stockton, S. et al. Comparative efficacy and acceptability of antimanic drugs in acute mania: a multiple-treatments meta-analysis. *The Lancet* 2011; 378(9799):1306-1315.

19. Buckley, P.F., Paulsson, B., Brecher, M. Treatment of agitation and aggression in bipolar mania: efficacy of quetiapine. *Journal of Affective Disorders* 2007; 100 Suppl 1:S33-S43.

20. Meehan, K., Zhang, F., David, S., Tohen, M., Janicak, P., Small, J. et al. A double-blind, randomized comparison of the efficacy and safety of intramuscular injections of olanzapine, lorazepam, or placebo in treating acutely agitated patients diagnosed with bipolar mania. *Journal of Clinical Psychopharmacology* 2001; 21(4):389-397.

21. Findling, R., Nyilas, M., Forbes, R., McQuade, R., Jin, N., Iwamoto, T. et al. Acute treatment of pediatric bipolar I disorder, manic or mixed episode, with aripiprazole: A randomized, double-blind, placebo-controlled study. *Journal of Clinical Psychiatry* 2009; 70(10):1441-1451.

22. Tohen, M., Kryzhanovskaya, L., Carlson, G., DelBello, M., Wozniak, J., Kowatch, R. et al. Olanzapine Versus Placebo in the Treatment of Adolescents With Bipolar Mania. *American Journal of Clinical Hypnosis* 2007; 164(10):1547-1556.

23. Roke, Y., van Harten, P., Boot, A., Buitelaar, J. Antipsychotic Medication in Children and Adolescents: A Descriptive Review of the Effects on Prolactin Level and Associated Side Effects. *Journal of Child & Adolescent Psychopharmacology* 2009; 19(4):403-414.

24. Correll, C.U. Assessing and maximizing the safety and tolerability of antipsychotics used in the treatment of children and adolescents. *The Journal of Clinical Psychiatry* 2008; 69:26.

25. Committee for proprietary mednicinal products (CPMP), European Medicines Agency. Note for guidance on clinical investigation of medicinal products for the treatment and prevention of bipolar disorder. 2001; available from http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003528.pdf

26. Committee for Medicinal Products for Human Use (CHMP), European Medicines Agency. European Public Assessment Report: Procedure No. EMEA/H/C/000471/II/0082 : Aripiprazole. 2013; available from http://www.ema.europa.eu/docs/en_GB/document_library/EPAR_-_Assessment_Report_-_Variation/human/000471/WC500138620.pdf (accessed Feb. 2013).

27. US Food and Drug Administration (FDA). Patent and Exclusivity Search Results from query on Appl No 021436 Product 001 in the OB_Rx list. 2013; available from http://www.accessdata.fda.gov/scripts/cder/ob/docs/patexclnew.cfm?Appl_No=021436&Product_No=001&table1=OB_Rx (accessed Feb. 2013).

28. National Institute for Health and Clinical Excellence. The management of bipolar disorder in adults, children and young people, in primary and secondary care: Draft for consultation. 2005; available from http://www.nice.org.uk/nicemedia/live/10989/30160/30160.pdf (accessed Feb. 2013).

29. DelBello, M., Schwiers, M., Rosenberg, H., Strakowski, S. A Double-Blind, Randomized, Placebo-Controlled Study of Quetiapine as Adjunctive Treatment for Adolescent Mania. *Journal of the American Academy of Child and Adolescent Psychiatry* 2002; 41(10):1216-1223.

30. Pavuluri, M., Henry, D., Carbray, J., Sampson, G., Naylor, M., Janicak, P. Open-label prospective trial of risperidone in combination with lithium or divalproex sodium in pediatric mania. *Journal of Affective Disorders* 2004; 82S:S103-S111.

31. Biederman, J., Mick, E., Hammerness, P., Harpold, T., Aleardi, M., Dougherty, M. et al. Open-Label, 8-Week Trial of Olanzapine and Risperidone for the Treatment of Bipolar Disorder in Preschool-Age Children. *Biological Psychiatry* 2005; 58:589-594.

32. Findling, R., McNamara, N., Youngstrom, E.A., Stansbrey, R., Gracious, B., Reed, M. et al. Double-Blind 18-Month Trial of Lithium Versus Divalproex Maintenance Treatment in Pediatric Bipolar Disorder. *Journal of the American Academy of Child and Adolescent Psychiatry* 2005; 44(5):409-417.

33. Search filters [Online] Scotland: Scottish Intercollegiate Guidelines Network; 2013. Available from http://www.sign.ac.uk/methodology/filters.html (accessed 2013 February 15). *Scottish Intercollegiate Guidelines Network* 2013.

34. Golder, S., McIntosh, H.M., Duffy, S., Glanville, J. Developing efficient search strategies to identify reports of adverse effects in MEDLINE and EMBASE. *Health Information and Libraries Journal* 2006; 23(1):3-12.

35. Findling, R.L., McNamara, N.K., Youngstrom, E.A., Stansbrey, R.J., Frazier, T.W., Lingler, J. et al. An open-label study of aripiprazole in children with a bipolar disorder. *Journal of Child and Adolescent Psychopharmacology* 2011; 21(4):345-351.

36. Metabolic Effects of Antipsychotics in Children. 2012; available from http://ClinicalTrials.gov/show/NCT00205699

37. Otsuka Pharmaceutical Development & Commercialization, I. Aripiprazole Open-Label, Safety and Tolerability Study. 2012; available from http://ClinicalTrials.gov/show/NCT00102518

38. Aripiprazole for the Treatment of Mania in Children and Adolescents With Bipolar Disorder. 2010; available from http://ClinicalTrials.gov/show/NCT00181779

39. Danielyan, A., Cannon, M., Kowatch, R.A. An open-label trial of aripiprazole monotherapy in children and adolescents with bipolar I disorder. *Biological Psychiatry* 2010; 67(9 SUPPL. 1):233S.

40. Biederman, J., Mick, E., Spencer, T., Doyle, R., Joshi, G., Hammerness, P. et al. An open-label trial of aripiprazole monotherapy in children and adolescents with bipolar disorder. *Cns Spectrums* 2007.

41. Ramos-Rios, R., Gago-Ageitos, A.M., Vidal-Millares, M., Mazaira-Castro, J.A., Dominguez-Santos, M.D. Clinical effects and tolerability of aripiprazole in children and adolescents with psychiatric disorders. *European Neuropsychopharmacology* 2009; 19:S691.

42. Long-term Tolerability and Safety of Aripiprazole in the Treatment of Pediatric Patients with Schizophrenia or Bipolar I Disorder. 2008.

43. Findling, R., Nyilas, M., Aurang, C., van Beek, A., Jin, N., Marcus, R. et al. Long term efficacy of aripiprazole in children (ages 10-17 years) with bipolar I disorder. *American College of Neuropsychopharmacology 46th Annual Meeting 2007* 2007.

44. Forbes, A. Long-term efficacy and safety of aripiprazole in children and adolescents with mania. *European Neuropsychopharmacology Conference: 21st ECNP Congress Barcelona Spain* 2008; 18(S4):S556-S557.

45. Loze, J., Forbes, R., Zhao, J., Chen, C., Carson, W., Iwamoto, T. et al. Line item analysis in child and adolescent patients with bipolar I disorder treated with Aripiprazole in an acute study (31-03-240). *European Child and Adolescent Psychiatry Conference: 14th International Congress of ESCAP European Society for Child and Adolescent Psychiatry Helsinki Finland 2011* 2011; 20:S185.

46. Loze, J., Mankoski, R., Zhao, J., Carson, W., Youngstrom, R., Findling, R. et al. Line item analysis in paediatric patients with bipolar I disorder treated with aripiprazole. *European Psychiatry Conference: 19th European Congress of Psychiatry, EPA 2011 Vienna Austria 2011* 2011; 26.

47. National Horizon Scanning Centre. Aripiprazole (Abilify) for children and adolescents with bipolar disorder (Structured abstract). *Birmingham: National Horizon Scanning Centre (NHSC)* 2008.

48. Pikalov, A., McQuade, R., Carson, W., Nyilas, M., Forbes, R., Iwamoto, T. et al. Somnolence and sedation in child and adolescent patients with bipolar I disorder treated with aripiprazole in an acute study with long term follow-up. *Bipolar Disorders Conference: 8th International Conference on Bipolar Disorder 2009* 2009; 11(S1):69.

49. Whitehead, R., Chen, C., Kim, E., Carson, W., Iwamoto, T., Mathew, S. et al. Patient assessed quality of life vs. clinician assessment in a trial of aripiprazole in pediatric

patients with bipolar disorder. *Value in Health Conference: ISPOR 14th Annual International Meeting 2009* 2009; 12(3):A173.

50. Whitehead, R., Chen, C., Kim, E., Carson, W., Iwamoto, T., Mathew, S. et al. Patient assessed quality of life vs. clinician assessment in a trial of aripiprazole in pediatric patients with bipolar disorder. *Bipolar Disorders 8th International Conference on Bipolar Disorder 2009* 2009; 11(S1):89.

51. Mankoski, R., Zhao, J., Carson, W., Mathew, S., Forbes, R. Young mania rating scale line item analysis in pediatric subjects with bipolar i disorder treated with aripiprazole in a short-term, double-blind, randomized study. *Journal of Child and Adolescent Psychopharmacology* 2011; 21(4):359-364.

52. Geller, B., Luby, J., Joshi, P., Wagner, K., Emslie, G., Walkup, J. et al. A Randomized Controlled Trial of Risperidone, Lithium, or Divalproex Sodium for Initial Treatment of Bipolar Disorder, Manic or Mixed Phase, in Children and Adolescents. *Archives of General Psychiatry* 2012; 69(5):515-528.

53. AstraZeneca. NDA 20-639/S-045 (acute schizophrenia) and S-046 (acute bipolar mania). Seroquel® (quetiapine fumarate) Tablets for the treatment of pediatric patients with schizophrenia (13-17 years old) or bipolar mania (10-17 years old). *Briefing Document for Psychopharmacologic Drugs Advisory CommitteeMeeting of June 9 and 10, 2009* 2009.

54. Haas, M., DelBello, M., Pandina, G., Kushner, S., Van Hove, I., Augustyns, I. et al. Risperidone for the treatment of acute mania in children and adolescents with bipolar disorder: a randomized, double-blind, placebo-controlled study. *Bipolar Disorders* 2009; 11:687-700.

55. Pavuluri, M., Henry, D., Findling, R.L., Parnes, S., Carbray, J., Mohammed, T. et al. Double-blind randomized trial of risperidone versus divalproex in pediatric bipolar disorder. *Bipolar Disorders* 2010; 12(6):593-605.

56. Committee for Medicinal Products for Human Use (CHMP), European Medicines Agency. Assessment Report For Abilify: Procedure No. EMEA/H/C/471/II/0039. 2009; available from http://www.ema.europa.eu/docs/en_GB/document_library/EPAR_-_Assessment_Report_-_Variation/human/000471/WC500020167.pdf (accessed Feb. 2013).

57. McClellan, J., Kowatch, R., Findling, R.L. Practice parameter for the assessment and treatment of children and adolescents with bipolar disorder. *J Am Acad Child Adolesc Psychiatry* 2007; 46(1):107-125.

58. Legislation.gov.uk. Mental Health Act 1983. 2013; available from http://www.legislation.gov.uk/ukpga/1983/20/contents (accessed Feb. 2013).

59. Baldessarini, R.J. Treatment research in bipolar disorder: issues and recommendations. *CNS Drugs* 2002; 16(11):721-729.

60. Licht, R.W. Drug treatment of mania: a critical review. *Acta Psychiatrica Scandinavica* 1998; 97(6):387-397.

61. Jing Zhang, DA. FDA Clinical Review: Abilify: Aripiprazole: NDA 21436/SE5-021. 2008; available from http://www.fda.gov/downloads/Drugs/DevelopmentApprovalProcess/DevelopmentResources/ucm071737.pdf (accessed Feb. 2013).

62. Cubells, J.F. A head-to-head comparison of risperidone and divalproex for treatment of pediatric bipolar disorder. *Current Psychiatry Reports* 2011; 13(2):82-83.

63. MacMillan, C.M., Withney, J.E., Korndorfer, S.R., Tilley, C.A., Mrakotsky, C., Gonzalez-Heydrich, J.M. Comparative clinical responses to risperidone and divalproex in patients with pediatric bipolar disorder. *Journal of Psychiatric Practice* 2008; 14(3):160-169.

64. Findling, R., Youngstrom, E., McNamara, N., Stansbrey, R., Wynbrandt, J., Adegbite, C. et al. Double-blind, randomized, placebo-controlled long-term maintenance study of aripiprazole in children with bipolar disorder. *The Journal of Clinical Psychiatry* 2012; 73(1):57-63.

65. National Institute for Health and Clinical Excellence. Guide to the methods of technology appraisal. 2008; available from http://www.nice.org.uk/media/b52/a7/tamethodsguideupdatedjune2008.pdf

66. Historic and projected mortality rates (qx) from the 2010-based UK life tables - Principal projection [http://www.ons.gov.uk/ons/publications/re-reference-tables.html?edition=tcm%3A77-257453 accessed May 2012]. *Office for National Statistics* 2010.

67. Laursen, T., Munk-Olsen, T., Nordentoft, M., Mortensen, P. Increased Mortality Among Patients smitted With Major Psychiatric Disorders: A Register-Based Study Comparing Mortality in Unipolar Depressive Disorder, Bipolar Affective Disorder, Schizoaffective Disorder, and Schizophrenia. *Journal of Clinical Psychiatry* 2007; 68(6):899-907.

68. 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. *Quality of Life Research* 2006; 15:1271-1280.

69. Revicki, D.A., Hanlon, J., Martin, S., Gyulai, L., Nassir, G.S., Lynch, F. et al. Patient-based utilities for bipolar disorder-related health states. *Journal of Affective Disorders* 2005; 87(2-3):203-210.

70. 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 and Quality of Life Outcomes* 2008; 6(105).

71. Boyle, S., Jones, G., Walters, S. Physical activity, quality of life, weight status and diet in adolescents. *Qual Life Res* 2010; 19(7):943-954.

72. Heeg, B., Buskens, E., Botteman, M., Caleo, S., Ingham, M., Damen, J. et al. The Cost-Effectiveness of Atypicals in the UK. *Value in Health* 2008; 11(7):1007-1021.

73. Ara, R., Brazier, J. Populating an economic model with health state utility values: moving towards better practice. *Health Economics and Decision Science Discussion Paper Series* 2010; 09/11.

74. Basu, S., Senthil, K.R., V. Childhood mood disorders: Myth or reality. *Indian Pediatrics* 2012; 49(1):14-15.

75. NHS Reference Costs 2010-2011. *Department of Health* 2011.

76. Curtis, L. Unit Costs of Health & Social Care 2011. *PSSRU* 2011.