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

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 available from the pivotal aripiprazole trial were: comorbid ADHD; 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 a cohort of overweight and obese children. 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 only presented the base case as an incremental analysis. The ERG presented the results for the sensitivity analysis 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

- 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' sensitivity analyses 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

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.

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. 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

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 did not explicitly include uncertainty in their multipliers and hence also in their assessment of uncertainty in the utility values. However, sensitivity analyses presented by the manufacturers in the clarification process showed that increasing utility values by 30% (with an upper limit of 1) or decreasing utility values by 30% did not have a large impact upon the conclusions.

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 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 76.

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

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.
Contrastingly the ERG notes that due to a lack of evidence, the manufacturers did not model the effects of prolactin increase and conservative assumptions were made regarding the incidences of EPS and somnolence for olanzapine. Furthermore, due to different results collection periods the adverse events rates were at four weeks for aripiprazole but at three weeks for the comparators [redacted] Each of these factors will disfavour aripiprazole.