in collaboration with:

ERASMUS UNIVERSITEIT ROTTERDAM
INSTITUTE OF HEALTH POLICY & MANAGEMENT

Maastricht University

Arsenic trioxide for treating acute promyelocytic leukaemia

Produced by Kleijnen Systematic Reviews Ltd. in collaboration with Erasmus University Rotterdam (EUR) and Maastricht University

Authors
Rob Riemsma, Reviews Manager, KSR Ltd, UK
Bram Ramaekers, Health Economist, Maastricht UMC
Sabine Grimm, Health Economist, Maastricht UMC
Debra Fayter, Systematic Reviewer, KSR Ltd, UK
Sohan Deshpande, Systematic Reviewer, KSR Ltd, UK
Piet Portegijs, Systematic Reviewer, KSR Ltd, UK
Nigel Armstrong, Health Economist, KSR Ltd, UK
Willem Witlox, Health Economist, Maastricht UMC
Xavier Pouwels, Health Economist, Maastricht UMC
Steven Duffy, Information Specialist, KSR Ltd, UK
Gill Worthy, Statistician, KSR Ltd, UK
Manuela Joore, Health Economist, Maastricht UMC
Jos Kleijnen, Director, KSR Ltd, Professor of Systematic Reviews in Health Care, Maastricht University

Correspondence to
Rob Riemsma, Kleijnen Systematic Reviews
Unit 6, Escrick Business Park
Riccall Road, Escrick
York, UK
YO19 6FD

Date completed 21/02/2018

Source of funding: This report was commissioned by the NIHR HTA Programme as project number STA 17/56/11.

Declared competing interests of the authors
None.

Acknowledgements
None

Commercial in confidence (CiC) data are highlighted in blue throughout the report.
Academic in confidence (AiC) data are highlighted in yellow throughout 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:
Riemsma R, Ramaekers B, Grimm S, Fayter D, Deshpande S, Portegijs, P, Armstrong N, Witlox W, Pouwels X, Duffy S, Worthy G, Joore M, Kleijnen J. Arsenic trioxide for treating acute promyelocytic leukaemia; A Single Technology Assessment. York: Kleijnen Systematic Reviews Ltd, 2018.

Contributions of authors
Rob Riemsma acted as project lead and systematic reviewer on this assessment, critiqued the clinical effectiveness methods and evidence and contributed to the writing of the report. Bram Ramaekers, acted as health economic project lead, critiqued the company's economic evaluation and contributed to the writing of the report. Sabine Grimm, Willem Witlox, Xavier Pouwels and Nigel Armstrong acted as health economists on this assessment, critiqued the company's economic evaluation and contributed to the writing of the report. Debra Fayter, Sohan Deshpande and Piet Portegijs acted as systematic reviewers, critiqued the clinical effectiveness methods and evidence and contributed to the writing of the report. Gill Worthy acted as statistician, critiqued the analyses in the company's submission and contributed to the writing of the report. Steven Duffy critiqued the search methods in the submission and contributed to the writing of the report. Manuela Joore critiqued the manufacturer's economic evaluation, contributed to the writing of the report and provided general guidance. Jos Kleijnen critiqued the company's definition of the decision problem and their description of the underlying health problem and current service provision, contributed to the writing of the report and supervised the project.

Abbreviations

6-MP	6-mercaptopurine
AATO	All-trans retinoic acid (ATRA) plus Arsenic trioxide (ATO)
AE	Adverse Events
AHSCT	Autologous haematopoietic stem cell transplantation
AML	Acute myeloid leukaemia
Amsa	Amsacrine
APL	Acute promyelocytic leukaemia
Ara-C	Cytarabine
ASH	American Society of Hematology (annual meeting)
ATO	Arsenic trioxide
ATRA	All-trans retinoic acid
BI	Budget impact
BIC	Bayesian information criterion
BNF	British National Formulary
BSC	Best supportive care
CDF	Cancer Drugs Fund
CE	Cost Effectiveness
CEA	Cost effectiveness Analysis
CEAC	Cost effectiveness Acceptability Curve
CHMP	Committee for Medicinal Products for Human Use
CHR	Complete Haematological Remission;
CI	Confidence Interval
CLL	Chronic lymphocytic leukaemia
CML	Chronic myeloid leukaemia
CMR	Complete Molecular Remission;
CNS	Central nervous system
CR	Complete response/remission
CRD	Centre for Reviews and Dissemination
CrI	Credible interval
CS	Company Submission
CSR	Clinical study report
CTCAE	Common Terminology Criteria for Adverse Events (NCI)
DFS	Disease-free survival
DNR	Daunorubicin
DT	Decision tree
DXM	Dexamethasone;
ECG	Electrocardiogram
ED	Early death
EFS	Event-free survival
EMA	European Medicines Agency
EORTC	European Organisation for Research and Treatment of Cancer
EPAR	European public assessment report
EQ-5D	European Quality of Life-5 Dimensions
ERG	Evidence Review Group
EUR	Erasmus University Rotterdam
FDA	Food and Drug Administration
FU	Follow-up
GO	Gemtuzumab ozogamicin
GvHD	Graft-versus-host disease

HADS	Hospital Anxiety and Depression Scale
HIV	Human immunodeficiency virus
HR	Hazard ratio
HRQL	Health-related Quality of Life
HSCT	Hematopoietic stem cell transplantation
HTA	Health Technology Assessment
ICER	Incremental cost-effectiveness ratio
ICUR	Incremental cost-utility ratio
IDA	Idarubicin
ITT	Intention to Treat
IV	Intravenous
KM	Kaplan–Meier
KSR	Kleijnen Systematic Reviews
L-VEF	Left-ventricular ejection fraction
LYG	Life Year Gained
MDS	Myelodysplastic syndrome
MeSH	Medical Subject Headings
MTC	Mixed Treatment Comparison
MTX	Methotrexate
MTZ	Mitoxantrone
NA	Not applicable
NCI	National Cancer Institute
NHS	National Health Services
NHSBT	National Health Service Blood and Transplant
NICE	National Institute for Health and Care Excellence
NIHR	National Institute for Health Research
NR	Not Reported
NS	Not significant
OS	Overall survival
PCR	Polymerase chain reaction
PRESS	Peer Review of Electronic Search Strategies
PSA	Probabilistic Sensitivity Analyses
PSS	Personal Social Services
QALY(s)	Quality-adjusted Life Year(s)
QLQ-C30	Quality of Life Questionnaire – Core 30
QoL	Quality of life
RFS	Relapse-free/Recurrence-free survival
RT-PCR	Reverse transcription polymerase chain reaction
RCT	Randomised Controlled Trial
SAE	Serious Adverse Events
SCT	Stem cell transplant
SD	Standard deviation
SEM	Standard error of the mean
SmPC	Summary of product characteristics
STA	Single Technology Appraisal
tMDS-AML	Therapy-related myelodysplastic syndrome or acute myeloid leukaemia
UK	United Kingdom
UMC	University Medical Centre
WBC	White blood cell
WHO	World Health Organisation

Table of Contents

Table of Tables

Table of Figures

1. SUMMARY

1.1 Critique of the decision problem in the company's submission

The NICE scope describes the decision problem as the clinical and cost effectiveness of arsenic trioxide (ATO) (with or without all-trans retinoic acid (ATRA)) within its marketing authorisation for adults with untreated low-to-intermediate risk acute promyelocytic leukaemia or relapsed/refractory acute promyelocytic leukaemia (APL).

The population in the submission is in line with the scope. Two main trials were included in the submission for patients with newly diagnosed APL (APL0406 and AML17). APL0406 took place in Italy and Germany whereas AML17 had trial centres in the UK, Denmark and New Zealand. The clinical expert from the company advised that "in the UK patients are treated following the AML17 protocol." However, AML17 also included patients at high risk who do not form part of the scope of this submission. In addition, in AML17 the dosing and regimens for ATO in the intervention arm (ATRA plus ATO (AATO)) were not in accordance with the licence; whilst the dosing and regimens in APL0406 were in accordance with the licence. As NICE can only issue guidance for interventions in accordance with the UK licence indication, APL0406 seems the most appropriate trial. However, AML17 might be a better reflection of UK practice.

The comparators listed in the NICE scope are: AIDA regimen (ATRA in combination with idarubicin), haematopoietic stem cell transplantation (HSCT) (for people with relapsed or refractory APL) and best supportive care (for people with relapsed or refractory APL). For first line treatment, AIDA was the comparator considered in the company submission (CS), both in the APL0406 trial and in the economic analysis. For adults with relapsed/refractory APL the company presented one randomised controlled trial (RCT) that included two arms: AATO versus ATO. Therefore, no comparative evidence for ATO in relation to any of the relevant comparators listed in the scope has been presented in the CS. Best supportive care and HSCT were not considered as comparators for people with relapsed or refractory APL in the submission.

1.2 Summary of clinical effectiveness evidence submitted by the company

The company presented evidence from three RCTs: Two of these were trials in newly diagnosed APL (APL0406 and AML17) and the third was a study in patients with relapsed APL (Raffoux, et al. 2003).

Newly diagnosed APL

Both trials in newly diagnosed APL (APL0406 and AML17) compared AATO (all-trans retinoic acid (ATRA) + ATO) with AIDA (ATRA + idarubicin). APL0406 included 266 patients with low-to-intermediate risk APL aged 18 to 71 years; while AML17 included 235 patients APL of any risk group, aged 16 or over (no upper age limit). APL0406 took place in Italy and Germany whereas AML17 had trial centres in the UK, Denmark and New Zealand. The dosing and regimens for the intervention arm (AATO) in AML17 were not in accordance with the licence; whilst the dosing and regimens for the intervention arm (AATO) in APL0406 were in accordance with the licence.

Results from APL0406 showed that AATO significantly improved overall survival (OS) at 50 months compared with AIDA (99.2% vs 92.6% respectively, p=0.007). The primary endpoint of this trial was event-free survival (EFS) at two years in an initial cohort of 156 patients (97% with AATO vs 86% with AIDA, p<0.001 for non-inferiority, p=0.02 for superiority). EFS was significantly better in the AATO group across all subsequent analyses to reach 97.3% at 50 months in the full cohort of 266 patients, compared with 80.0% in the AIDA group (p<0.001). The primary source of the observed EFS benefit was a reduction in the number of relapses with AATO – at 50 months, the cumulative incidence

of relapse was 1.9% in the AATO group compared with 13.9% in the AIDA group (p=0.0013). In terms of adverse events, corrected QT interval (QTc) prolongation was more common in the AATO group in the induction phase of treatment (8.5% vs 0.7%); as was grade 3 to 4 hepatic toxicity (40% vs 3%). However, there were no significant differences between groups in numbers of patients with moderate to severe differentiation syndrome in induction. During all treatment phases there were 19 instances of neurotoxicity with AATO and 0 with AIDA. In the AATO group patients experienced fewer haematological adverse events including fever and infection episodes and fewer grade 3 to 4 neutropenia and thrombocytopenia lasting over 15 days.

Results from AML17 showed an EFS benefit of AATO over AIDA (four-year EFS of 91% vs 70%, p=0.002), particularly in low-risk patients (four-year EFS was 92% in the AATO group [n=86] vs 71% in the AIDA group [n=92], p=0.008). The four-year cumulative incidence of haematological relapse was 18% in the AIDA arm and 1% in the AATO arm (p=0.0007). In this trial, patients were closely monitored for molecular relapse and many were treated before progression into a full haematological relapse, so that the cumulative incidence of molecular relapse at four years was 27% in the AIDA group and 0% in the AATO group (p<0.0001).

Relapsed or refractory APL

The study by Raffoux et al. (2003) compared AATO with ATO, which is not a relevant comparison according to the NICE scope. OS was similar between the AATO and ATO study arms. Across both groups, the estimated two-year OS was 59% (95% CI: 35%–77%). EFS was not reported in this study.

EMA approval of ATO in patients with relapsed or refractory APL was based on two single-arm studies conducted in the US, with no additional European studies supporting the EMA approval in this indication. However, these two studies were not included in the company submission.

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

The company conducted systematic reviews of the evidence for arsenic trioxide and its comparators in newly diagnosed and relapsed/refractory patients as per the NICE scope. The submission and response to clarification provided sufficient details for the ERG to appraise the literature searches. A good range of databases were searched, and additional searches of conference proceedings were conducted. Searches were carried out in accordance with the NICE guide to the methods of technology appraisal sections 5.2.2 and 5.2.4.

Of the two trials presented as evidence for untreated APL (APL0406 and AML17) only one (APL0406) is in accordance with the licence. We have thus prioritised an assessment of this trial in our report and presented AML17 as supporting evidence only. There are further differences between the trials which are outlined in this report. A full assessment of the quality of APL0406 by the company and by the ERG is hampered by the fact that only published information is available for assessment as the trial was not conducted by Teva. Overall the trial appears to have been well conducted. It is important to note that there are no UK patients in APL0406. The committee will need to consider the importance of this issue given that the population (low and intermediate risk), the intervention and the comparator are relevant to the UK setting. The effectiveness data show that relevant patient outcomes are improved. The safety data show that patients will need to be carefully selected and informed of the particular risks of the chosen regimen. Knowledge of long-term toxicity of AATO for newly diagnosed patients awaits a post-authorisation long term safety cohort study.

The company presented one trial in relapsed/refractory patients. The trial by Raffoux et al. (2003) compared AATO with ATO, which is not a relevant comparison according to the NICE scope. We have

not reported in detail on this small trial. In view of this lack of relevant evidence, the ERG considers that non-RCTs could have been included in the submission for the relapsed/refractory population. The committee will need to consider whether it is necessary to explore the evidence further given the company's view that "the use of ATO in the relapsed or refractory APL setting is already so well-established in routine clinical practice that it would be difficult to provide NICE with novel information based on the analysis of additional studies."

No trials of ATO alone were presented for those with relapsed/refractory disease. The committee will need to decide if they are in agreement with the company that ATO alone is rarely used in UK practice. It should also be noted that no trials in the CS compared ATO regimes with hematopoietic stem cell transplantation or with best supportive care as specified in the NICE scope.

1.4 Summary of cost effectiveness evidence submitted by the company

The company conducted systematic literature reviews (SLRs) to identify relevant cost effectiveness studies, health-related quality of life studies, resources and costs studies. Although the SLR identified cost effectiveness analyses (CEAs) in the literature, the company decided to develop a de novo model. The model structure proposed by the company however diverges from the one used in the CEAs identified in the SLR. The company justified this by stating that the existing economic evaluations did "not adequately reflect the trajectory of APL patients" and hence developed a more complex model structure to "offer more granularity with treatment phases, molecular remission and HSCT" and better reflect the clinical trajectory of APL patients. The model structure developed by the company considered different treatment phases: first line, second line, hematopoietic stem cell transplantation (HSCT) (including both alloHSCT and autoHSCT) and other phases (i.e. treatment-related myelodysplastic syndrome or acute myeloid leukaemia (tMDS/AML) and death).

The model adopts the perspective of the NHS and Personal and Social Services (PSS) in England and Wales. The model time horizon is 40 years, at the end of which a significant proportion of patients in the model are still alive (>40% of patients in the ATRA+ATO first line and AIDA second line arm). The model cycle length is four weeks to capture the treatment schedule and a half-cycle correction is applied. All costs and health gains were discounted at a rate of 3.5% per year.

The company only assessed the cost effectiveness of ATRA+ATO (AATO) in the newly diagnosed low-to-intermediate risk APL population, i.e. in first line treatment. The cost effectiveness of AATO in the relapsed/refractory APL population was not assessed.

AATO was modelled with up to two cycles (of four weeks) of induction therapy followed by eight cycles (of four weeks) of consolidation therapy. The only comparator, first line AIDA, was implemented with up to two cycles (of four weeks) of induction therapy followed by three cycles (of four weeks) of consolidation therapy. For both AATO and AIDA, maintenance treatment was not modelled and the justification provided by the company was that it is usually omitted in UK clinical practice with the aim of minimising the risk of tMDS/AML.

The transition probabilities from the first line phase of the model were informed by the APL0406 trial. The transitions from second line states and the HSCT states were only sparsely described.

Both the APL0406 and the AML17 trials used the EORTC QLQ-C30 instrument, and not the EQ-5D, to measure health-related quality of life (HRQoL) outcomes. Hence, utility values were obtained from the literature. However, no study reporting utility values based on the EQ-5D for APL patients was identified in the literature. Instead, utilities obtained in other diseases (e.g. chronic lymphocytic leukaemia and acute myeloid leukaemia) were used as a proxy for APL utilities. Additionally, the

company performed multiple adjustments to these utilities with the intention to make them more relevant for the modelled population.

The cost categories included in the model were treatment acquisition costs, medical costs (treatment administration, supportive care, monitoring and follow-up, HSCT, palliative care), and costs of managing adverse events. Drug costs were based on the British National Formulary (BNF) while NHS reference costs, BNF and PSSRU were mainly used for the medical costs. NHS reference costs were used to inform the costs of managing adverse events; alternatively, published literature was used.

In the company base-case (probabilistic) AATO was less expensive (£31,088 saved) and more effective (2.546 QALYs gained) than AIDA and thus the dominating strategy for newly diagnosed low-to-intermediate risk APL (i.e. the first line population). The probability of AATO being cost effective at a willingness-to-pay (WTP) of £30,000 per QALY was 94%. AATO remained dominant in most of the sensitivity and scenario analyses conducted by the company.

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

The cost effectiveness searches in the company submission were reported in enough detail for the ERG to appraise them. Separate searches were conducted to identify cost effectiveness studies, health-related quality of life data, and cost and healthcare resource use data.

The ERG considers that, although it is more complex than published cost effectiveness studies, the model structure is appropriate to reflect this condition and treatment pathway. The main ERG concerns regarding the model structure relate to inconsistencies between treatments regarding the modelling of patients that cannot be evaluated for molecular remission, an error in the number of tunnels used to represent the two year molecular remission health state, the absence of disease-related mortality from on treatment health states and the applicability of alloHSCT to the UK clinical setting. These issues were considered in the additional analyses performed by the ERG.

The model time horizon of 40 years results in a significant proportion of patients alive at the end of the model time horizon. Hence, the time horizon was extended to 56 years in the ERG base-case.

AATO was only assessed in the newly diagnosed APL population (first line). Although, in its clarification response, the company provided an analysis in the relapsed/refractory population (second line), the company's description of this analysis did not provide clarity over precisely how this analysis was performed. The ERG therefore implemented their own scenario by removing the first line health states and using the second line transition probabilities to reflect the relapsed/refractory population.

Inconsistent with the scope, the company did not consider ATO stand-alone nor best supportive care (BSC) as comparators in the second line setting. However, the ERG believed the justifications from the company to exclude these comparators, highlighting (based on expert opinion) that ATO alone and BSC would only rarely be used in UK clinical practice in the second line setting, to be reasonable.

The ERG had multiple concerns related to the estimation of treatment effectiveness. This included multiple reference/calculation errors, the overestimation of cardiac events and thus patients switching to second line induction for AIDA, assumptions and calculation errors related to the relapse probabilities and not considering treatment switching due to reversible arrhythmia in the model. Additionally, the evidence to inform transitions from second line health states was weak and it was frequently not transparently reported how the transition probabilities were obtained. Similarly, most of the evidence sources to inform transition probabilities from the HSCT health states are not described in the CS (neither are the transition probabilities reported). The lack of detailed description and

justification is worrying, given treatment effectiveness (including implicit assumptions made and selection of evidence sources to obtain transition probabilities) is often an influential part of the cost effectiveness model. This includes assumptions regarding the extrapolation of treatment effectiveness which is not extensively discussed in the CS. These issues were considered in the additional analyses performed by the ERG.

The ERG agrees with the company that utility values for APL patients elicited through the EQ-5D are probably not available in the literature. However, the ERG is concerned with the validity of the utility values for the following reasons: the selection process of the utility values and the assumptions underlying disutilities associated with adverse events were unclear, the non-adherence with the NICE reference case, and the lack of justification supporting the adjustments made by the company. The ERG preferred not to use the company's adjustment in its base-case analysis, and instead used the unadjusted health state utilities. Additionally, in order to prevent health state utility values exceeding the general population utility values (over time), the ERG decided to cap the health state utility values in the model using the general population utility values.

The main concerns regarding resource use and costs in the model relate to the lack of justification regarding some of the sources used. The ERG asked the company to provide more specific justification for each resource use and cost item. The company responded that they aimed to use NHS reference costs and the PSSRU wherever possible, supplementing this with data from studies identified through a targeted search where necessary. However, the company did not provide further justification and details about the included targeted sources, and the ERG was therefore unable to assess whether these sources were the best available evidence to inform resource use and costs estimates.

Considering the validity of the cost effectiveness results presented by the company, the ERG perceives the expected life expectancy outcome of the model to be relatively long. This is likely linked to the lack of disease-related mortality in the model during the first line and second line health states (only general population mortality is considered) as well as assumptions concerning (the extrapolation of) treatment benefits. The undiscounted life years (LYs) and QALYs for AATO, estimated in the model, are 33.22 and 27.91 respectively. When extending the model time horizon to 56 years, to represent a life time horizon, which is consistent with the NICE reference case, these increase to 35.83 and 30.12 respectively. The ERG is uncertain whether these outcomes have face validity. Particularly given that in the general UK population, the LY and QALYs estimated for patients aged 45 (with 48.7% being male) are 37.62 and 29.62 respectively.

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

1.6.1 Strengths

Overall, the company submission searches were well presented and reproducible. Searches were carried out on a range of databases and supplementary resources. The clinical evidence for untreated patients is based on a randomised controlled trial which is relevant to the population in this appraisal.

Strengths related to the economic evaluation include the granularity the model structure provides in comparison with other CEAs identified in the SLR. However, related to this, the (lack of) data to inform post first line transition probabilities can be regarded as a limitation. Additionally, the lack of (EQ5D) utility values for the APL population is a concern. Nevertheless, it is reassuring that AATO for the first line population remained dominant in the ERG base-case, and that the worst-case scenario produced by the ERG resulted in an (deterministic) ICER of £21,622 per QALY gained.

1.6.2 Weaknesses and areas of uncertainty

The ERG was concerned about the overall quality of the searches conducted, as there were numerous inconsistencies, inaccuracies and redundancy throughout. It is possible that relevant evidence may have been missed. However, the main weakness of the submission is that only one trial is directly relevant to the appraisal (APL0406) which provides data on an untreated population only. The trial does not have any UK patients. The company presented one trial in relapsed/refractory patients. However, the trial did not present a relevant comparison according to the NICE scope. The committee will need to consider whether it is necessary to explore further the evidence for relapsed/refractory patients or whether it is sufficiently well-established in routine clinical practice.

Although decision uncertainty in the economic evaluation is relatively low, suggested research priorities regarding the cost effectiveness might be focused on obtaining health state utility values for the APL population as well as transition probabilities from and to the HSCT health states reflective of UK clinical practice.

1.7 Summary of exploratory and sensitivity analyses undertaken by the ERG

In the company base-case (probabilistic) AATO is less expensive (£31,088 saved) and more effective (2.546 QALYs gained) than AIDA and thus the dominating strategy for newly diagnosed low-to-intermediate risk APL (i.e. the first line population). AATO remained dominant in most of the sensitivity and scenario analyses conducted by the company. The ERG has incorporated various adjustments to the company base-case. This resulted in the (deterministic) ERG base-case, wherein AATO remained dominant. Moreover, the ERG produced a worst-case scenario (combination of some of the scenario analyses explored by the ERG), to acknowledge the uncertainties discussed by the ERG in this report. This resulted in an ICER of £21,622 per QALY gained (deterministic). The ERG was unable to perform probabilistic analysis for its base-case. However, the ERG does not consider this to be a major issue as AATO is likely to remain dominant if the ERG would be able to produce probabilistic results for its base-case.

In conclusion, despite the ERG's criticism of the economic model and several highlighted uncertainties, it is reassuring that AATO for the first line population remained dominant in the ERG base-case, and that the worst-case scenario produced by the ERG resulted in an ICER of £21,622. However, as indicated by the subgroup analysis performed by the ERG, the cost effectiveness of AATO for the second line might be substantially different (estimated ICER of £31,184 per QALY gained).

2. BACKGROUND

In this report the ERG provides a review of the evidence submitted by Teva in support of arsenic trioxide, trade name TRISENOX®, for the treatment of patients with acute promyelocytic leukaemia (APL). In this section we outline and critique the company's description of the underlying health problem and the overview of current service provision. The information is taken from Chapter 3 of the company submission (CS) with sections referenced as appropriate.

2.1 Critique of company's description of underlying health problem.

The underlying health problem of this appraisal is APL which is a distinct subtype of acute myeloid leukaemia (AML). The company describes APL as a rare disease which is "caused by a translocation between chromosomes 15 and 17, abbreviated as t(15;17), fusing the PML gene with the RARA gene, which results in formation of the PML-RARα fusion protein".[1]

According to the NICE scope, "there were 2,590 diagnoses of acute myeloid leukaemia and 2,127 deaths in England in 2014. Around 10% of AML cases are APL.[2] The CS states that the exact incidence estimates vary across reports for example. Sant et al. analysed 2000 to 2002 data from 44 cancer registries across Europe and reported an overall annual crude incidence rate of 0.14 per 100,000[3]; Visser et al. analysed 1995 and 2002 data from 64 European cancer registries reporting a crude annual incidence rate of 0.11 per 100,000 people;[4] Dores et al. conducted a study based on the US Surveillance, Epidemiology and End Results (SEER) Program registry reported that age-adjusted incidence of APL was 0.27 per 100,000 person-years.[5]

The CS states that the "age distribution is a key difference between APL and most other AML types, which are diagnosed at a median age exceeding 60 years".[1] Hence, APL is likely to pose a considerable societal burden, affecting people of working age.[1]

The CS states that APL can progress rapidly with very poor survival prognosis. The company mentions a retrospective analysis that reported about 10–29% of patients die within 30 days of hospital admission or diagnosis. The majority of these deaths are due to haemorrhage (CNS or pulmonary) (31 to 55%) because of high risk of coagulopathy in APL patients.[6]

The CS refers to relapse risk stratification which is used to determine the most appropriate treatment options for APL patients.[1] The CS states that "assessment of relapse risk in APL is primarily based on white blood cell (WBC) count at presentation, with patients whose WBC count exceeds 10×10^9/L generally predicted to have a higher risk of relapse. Risk stratification was developed through a joint analysis of two multicentre trials (AIDA0493 and LPA96)".[7] The relapse risk categories are: low-risk (WBC $\leq10\times10^9$/L and platelet count $>40\times10^9$/L), intermediate-risk (WBC and platelet counts $\leq10\times10^9$/L and $\leq40\times10^9$/L, respectively) and high-risk (WBC count $>10\times10^9$/L).[7] In this submission, the population under consideration is adults with untreated low-to-intermediate risk and relapsed/refractory APL.

ERG comment:

- The company provides a good overview of the underlying health problem. The ERG checked the references provided to support the statements in the company submission. In general, these were found to be appropriate.

2.2 Critique of company's overview of current service provision

The company correctly reports that there is no relevant technology appraisal guidance on APL published in the UK to date and that ATO has never been assessed by the NICE. The CS mentions that, in the past

patients with newly-diagnosed APL were commonly treated with the standard chemotherapy-based treatment approach, AIDA which is a combination of all-trans retinoic acid (ATRA) and idarubicin.[1] In 2015, Teva conducted primary market research in seven European countries including Austria, France, Germany, Italy, Spain, Switzerland and the UK to understand the current treatment patterns in APL.[1] The CS states that "patients newly-diagnosed with APL in the UK are commonly treated according to clinical trial protocols (MRC AML trials), as they are recommended to enrol in ongoing trials upon diagnosis".[1]

Arsenic trioxide has a UK marketing authorisation for induction and consolidation in adult patients with: newly diagnosed low-to-intermediate risk APL (white blood cell count, $\leq$ 10 x $10^3/\mu l$) in combination with ATRA and relapsed/refractory APL (previous treatment should have included a retinoid and chemotherapy).[2]

The CS states "first-line therapy in APL generally consists of three consecutive treatment phases: induction, consolidation and maintenance, although maintenance is usually omitted in the UK clinical practice with the aim of minimising the risk of treatment-related myelodysplastic syndrome or acute myeloid leukaemia (tMDS/AML)"[1] The ATO-based first-line treatment regimen was not explicitly recommended for wider use by the 2013 guidelines from the European Society for Medical Oncology (ESMO)[8] and the 2009 European LeukemiaNet guidelines.[9] However, the two German guidelines Deutsche Gesellschaft für Hämatologie und Medizinische Onkologie (DGHO)[10] and the German Intergroup[11] have listed the AATO (ATRA+ATO) combination as an option for treating newly-diagnosed low-to-intermediate-risk patients.

The CS states that "patients with relapsed or refractory APL may receive a HSCT to consolidate second remission" if considered at risk of additional relapses.[1]However, patients who are not transplant candidates may receive additional ATO cycles.[1]

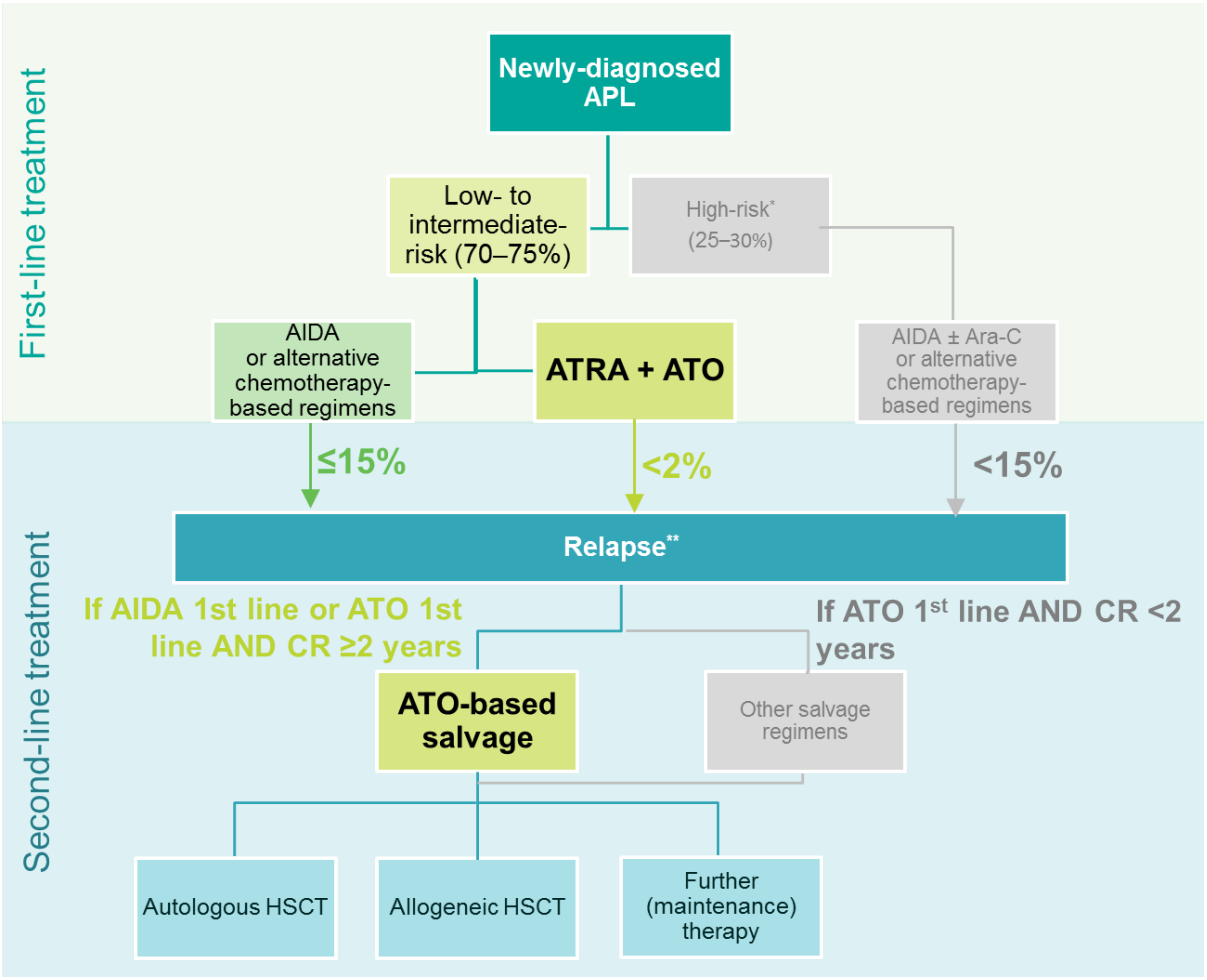
The CS states that "according to expert opinion, patients in the UK are treated as soon as molecular relapse is detected and before the patient progresses into a haematological relapse."[1] Once a relapse is confirmed, the choice of second line treatment depends on the type of first line therapy the patient has received for e.g. UK patients could switch from AIDA to AATO (ATRA+ATO) and from AATO to AIDA.[1]

The CS highlights the fact that "there is a lack of well-established paradigms or guidelines for second-line treatment following AATO administration in first line, and the field is constantly evolving with growing experience of first-line ATO use".[1] Hence, based on expert opinion the economic analysis in this submission included mixed re-treatment/switch approach which assumes that the "patients who remained in remission for 2 years or longer following first-line AATO treatment were re-treated with AATO upon relapse, while patients who achieved only a short (<2 years) remission after first-line treatment with AATO were treated with AIDA".[1]

The CS states that in "UK patients, second remissions are often consolidated with a HSCT". Also, according to the clinical expert "allogeneic HSCT is generally used in patients who enter haematological remission following second-line treatment but fail to achieve molecular remission; in patients who achieve a second molecular remission, allogeneic HSCT is rarely considered due its associated risks."[1] Further, the company's clinical expert also suggested that the "patients salvaged with ATO do not necessarily need transplantation, while those salvaged with chemotherapy generally do."[1]

Figure 2.1 shows the perceived role of ATO in the treatment of both newly diagnosed and relapsed/refractory patients with APL

Figure 2.1: Simplified treatment pathway in APL showing the licensed indications for ATO



Source: Section B1.3.2.2 of the CS
AIDA = ATRA in combination with idarubicin; APL = acute promyelocytic leukaemia; ATO = arsenic trioxide; ATRA = all-trans retinoic acid; CR = Complete remission; HSCT = haematopoietic stem cell transplantation

ERG comment:

- The company's overview of current service provision is appropriate and relevant to the decision problem under consideration.

3. CRITIQUE OF COMPANY'S DEFINITION OF DECISION PROBLEM

Table 3.1: Statement of the decision problem (as presented by the company)

	Final scope issued by NICE	Decision problem addressed in the company submission and rationale	ERG comments
Population	Adults with: • untreated low-to-intermediate risk acute promyelocytic leukaemia • relapsed/refractory acute promyelocytic leukaemia (APL)	Adults with: • untreated low-to-intermediate risk acute promyelocytic leukaemia • relapsed/refractory acute promyelocytic leukaemia (APL) characterised by the presence of the t(15;17) translocation and/or the presence of the promyelocytic leukaemia/ retinoic-acid-receptor-alpha (PML/RAR-alpha) gene.	In line with the scope.
Intervention	ATO (with or without ATRA)	• First line treatment: ATO combined with ATRA; both administered according to the APL0406[12] protocol. AML17[13] protocol was studied as a scenario. • Second line treatment: ATO administered according to the SmPC + ATRA administered according to the APL0406[12] protocol (as in first line). The AML17 protocol[13] was studied in a scenario analysis. Rationale: In line with both the pivotal APL0406 trial[12, 14] and the AML17 trial[13], ATO is authorised for use in newly-diagnosed patients in combination with ATRA. No treatment combinations are specified for use in relapsed/refractory patients, although in the AML17 trial treatment with ATRA+ATO (administered as in first line) was used in patients who relapsed.[15] Based on clinical expert opinion, it appears ATO alone (without ATRA) is now rarely used in the relapsed/refractory setting. Thus, for both first- and second-line treatment, only the ATRA+ATO combination was considered in the economic analysis.	The company presented evidence for AATO (ATRA+ATO) only. No evidence was presented for ATO alone.
Comparator(s)	• AIDA regimen (ATRA in combination with idarubicin) • haematopoietic stem cell transplantation (HSCT)	• Following a relapse, the choice of therapy strongly depends on prior treatments the patient has received. It is therefore difficult to separate first- and second-line indications of ATO, as they're closely linked. To optimally reflect the treatment pathway of APL patients in the UK, Teva has decided to submit a single model which evaluates the cost-	For first line treatment, one trial is presented comparing AATO versus AIDA (the APL0406 trial[12, 14]). A second trial was presented (the AML17 trial);[13]

	Final scope issued by NICE	Decision problem addressed in the company submission and rationale	ERG comments
	(people with relapsed or refractory APL) • best supportive care (people with relapsed or refractory APL)	effectiveness of ATO (+ATRA) in newly-diagnosed patients (first line indication) with second line treatments included, rather than presenting a separate cost effectiveness evaluation of ATO as a second line treatment. • For first line treatment, AIDA was the comparator considered in both the pivotal APL0406 trial[12, 14] and in the economic analysis • For the second-line part of the model, we considered a situation where ATO was available first-line and some of the patents who received ATO first line switched to AIDA in second line, so that AIDA was retained as the comparator. Rationale: • In the second line indication, HSCT was not considered as a direct comparator, since administration of ATRA+ATO usually precedes transplantation rather than replaces it. Upon relapse, ATRA+ATO can be used to induce remission, which, if possible, would be consolidated with HSCT.[11, 16] Although additional ATO (+ ATRA) cycles may be used in patients who do not undergo a transplant,[16, 17] ATO-based maintenance treatment is not included in the licensed administration schedule, and was therefore not considered in the economic analysis. Furthermore, other maintenance treatment options are also available to APL patients who do not undergo transplantation,[16] and it would be difficult to include all of them without overtly complicating the analysis. We therefore took a simplified approach of not modelling second line maintenance treatment, especially given that the number of patients concerned would be very small. • Best supportive care was not considered as a direct comparator in the second line indication. Following ATO-based treatment of first APL relapse, Lengfelder et al. reported 3-year EFS of ≥45%,[17] suggesting that attempting curative treatment may be most appropriate in patients with relapsed/refractory APL. Given the severity of APL, best supportive care can be seen as a palliative approach, and thus expected to be used where the disease is refractory to all other treatments, including ATO in second (or subsequent) treatment lines. Thus, it is unlikely that best	however, in this trial ATO was not administered according to its licensed indication). For second line treatment, a trial is presented comparing AATO vs ATO (Raffoux et al. 2003[18]). As ATO is part of both arms, this trial is not informative for the effectiveness of ATO in second line. No evidence is presented for HSCT and best supportive care. The company did not consider the relapsed/refractory APL population neither did they consider BSC nor ATO alone in the health economic sections of the CS. This is discussed in more detail in sections 5.2.4 and 5.2.3 of this report.

	Final scope issued by NICE	Decision problem addressed in the company submission and rationale	ERG comments
		supportive care will be considered an alternative to ATO or AIDA (see below) for treatment of relapsed APL. It is, however, worth noting that the economic analysis does take into account best supportive care – upon failure of second line treatment, patients in the model progressed to an end-of-life state, where they received palliative care. • The choice of second line treatment is largely determined by the first line therapy that the patient has received, and ATO (usually + ATRA) is the standard treatment for APL relapses after first line treatment containing ATRA and an anthracycline (e.g. AIDA). However, the choice of optimal salvage treatment in patients who relapse following first line ATO use is less clear. This is largely due to the absence of established guidelines, as many treatment guidelines in APL (e.g. from the European LeukemiaNet[9] and ESMO[8]) precede the approval of ATO for first-line use. In the economic analysis, treatment of relapses following first line ATO use was therefore based on clinical expert opinion. It was assumed that patients who remained in remission for ≥2 years following first line ATRA+ATO treatment were re-treated with ATRA+ATO upon relapse. However, patients who achieved only a short (<2 years) remission after first line treatment with ATRA+ATO, were assumed to be treated with AIDA upon relapse. Thus, AIDA was considered as a comparator also in the relapsed/refractory APL setting.	
Outcomes	• Overall survival (OS) • Progression-free survival (PFS) • Response rates (bone marrow remission) • Adverse effects of treatment • Health-related quality of life (HRQoL)	• OS • Event-free survival (EFS) • Complete haematological and molecular remission rates • Cumulative incidence of relapse (CIR) • Disease-free survival (DFS) or relapse-free survival (RFS) • Adverse effects of treatment • HRQoL Rationale: • PFS was not an endpoint in the pivotal APL0406 trial[12, 14] or in the AML17 trial,[13] and is thus not presented. Instead, the manufacturer	EFS was used instead of PFS. In the main trial, APL0406, EFS was assessed at 2 years after diagnosis, with treatment failure defined as any of the following: 1) no achievement of hematologic complete remission (CR) after induction; 2) no achievement of molecular CR after three consolidation courses; 3) molecular relapse;

	Final scope issued by NICE	Decision problem addressed in the company submission and rationale	ERG comments
		presented data on EFS – the primary endpoint of the APL0406 trial.[12, 14] It is, however, worth noting that in the APL0406 trial patients failing treatment were those who did not achieve remission, relapsed, or died, which is similar to what would be considered treatment failure when analysing PFS. In the AML17 trial, an additional event of treatment-related myelodysplastic syndrome or acute myeloid leukaemia (tMDS/AML) was also included in the EFS analysis; however, only a single patient in this study developed tAML,[13] so that inclusion of this event in EFS evaluation could be considered to have little effect on the overall result. In conclusion, although EFS rather than PFS is presented, the two outcomes are similar, so this does not represent a major deviation from the scope. • In addition to the outcomes listed in the Final Scope, the manufacturer will also present data on cumulative incidence of relapse and DFS (or RFS), if available. Given the curative intent of APL treatment, these endpoints are of particular importance, as they provide information on the proportion of patients who remain disease-free.	4) haematological relapse, or 5) death. EFS in this case is similar to PFS.

Source: Table 1.1, Section B.1.1 of the CS

AATO = ATRA+ATO; AIDA = ATRA in combination with idarubicin; AML = acute myeloid leukaemia; APL = acute promyelocytic leukaemia; ATO = arsenic trioxide; ATRA = all-trans retinoic acid; CIR = cumulative incidence of relapse; CR = Complete remission; CS = company submission; DFS = disease-free survival; EFS = event-free survival; ERG = Evidence Review Group; HRQoL = health-related quality of life; HSCT = haematopoietic stem cell transplantation; OS = overall survival; PFS = progression-free survival; PML = promyelocytic leukaemia; RAR-alpha = retinoic-acid-receptor-alpha; RFS = relapse-free survival; SmPC = Summary of product characteristics; tMDS = treatment-related myelodysplastic syndrome

3.1 Population

The population defined in the scope is adults with untreated low-to-intermediate risk acute promyelocytic leukaemia (APL) and adults with relapsed/refractory APL. The population in the submission is in line with the scope.

Two main trials were included in the submission for patients with newly diagnosed APL (APL0406 and AML17). APL0406 took place in Italy and Germany whereas AML17 had trial centres in the UK, Denmark and New Zealand. The clinical expert from the company advised that "in the UK patients are treated following the AML17 protocol."[19] However, AML17 also included patients at high risk who do not form part of the scope of this submission and the dosing and regimens for the intervention arm (AATO) in AML17 were not in accordance with the licence; while the dosing and regimens for the intervention arm (AATO) in APL0406 were in accordance with the licence.

APL0406 seems the most appropriate study as NICE can only issue guidance for interventions in accordance with the UK licence indication. However, AML17 might be a better reflection of UK practice.

3.2 Intervention

The intervention (ATO with or without ATRA) is in line with the scope. Regulatory approval by the EMA for the treatment of relapsed or refractory patients was granted in 2002. In November 2016 it was approved in the EU for the treatment of newly-diagnosed patients with low-to-intermediate risk APL.[20]

ATO is indicated for induction of remission, and consolidation in adult patients with:

- Newly diagnosed low-to-intermediate risk acute promyelocytic leukaemia (APL) (white blood cell count, $\leq 10 \times 10^3/\mu l$) in combination with all-trans-retinoic acid (ATRA)
- Relapsed/refractory APL (Previous treatment should have included a retinoid and chemotherapy) characterised by the presence of the t(15;17) translocation and/or the presence of the Pro-Myelocytic Leukaemia/Retinoic-Acid-Receptor-alpha (PML/RAR-alpha) gene.

ATO must be administered under the supervision of a physician who is experienced in the management of acute leukaemias and special monitoring procedures apply (see CS, Table 1.2, pages 9-11).

ATO is indicated for the treatment of APL characterised by the presence of the t(15;17) translocation and/or the presence of the Promyelocytic Leukaemia/Retinoic-Acid-Receptor-alpha (PML/RAR-alpha) gene. This translocation accounts for up to 98% of APL cases; however, other translocations involving the *RARA* gene have also been identified in APL.[21] It is widely accepted that the diagnosis of APL (as opposed to other types of AML) should be confirmed through molecular testing for PML-RARA. Although the pivotal APL0406 trial accepted a number of methods through which genetic confirmation of APL diagnosis could be established,[12] the diagnostic tests that appear most feasible for routine use are polymerase chain reaction (PCR) and fluorescent *in situ* hybridisation (FISH).[1]

APL patients also undergo repeated bone marrow biopsies and the collected material is PCR-tested for the presence of PML-RARA, which allows the treating clinician to establish how the patient responds to treatment (i.e. if molecular remission has been achieved or if minimal residual disease can be detected), and to monitor the patient for molecular relapse (i.e. the reappearance of PML-RARA in the bone marrow), which allows second line treatment to be administered early, before the patient progresses into a full haematological relapse that may be life-threatening. The frequency of monitoring depends on treatment choice.

3.3 Comparators

The description of the comparators in the NICE scope is as follows:

- AIDA regimen (ATRA in combination with idarubicin)
- haematopoietic stem cell transplantation (HSCT) (people with relapsed or refractory APL)
- best supportive care (people with relapsed or refractory APL)

For first line treatment, AIDA was the comparator considered in the CS, both in the APL0406 trial[12, 14] and in the economic analysis.

For adults with relapsed/refractory APL the company presented one randomised controlled trial (RCT) that included two arms: AATO versus ATO. Therefore, no evidence for ATO in relation to any of the relevant comparators listed in the scope has been presented in the CS. The company justifies this by stating: "This was motivated by the well-established and widespread use of ATO in relapsed/refractory APL, and the fact it has long been considered first-choice therapy for induction and consolidation in this setting." (CS, section B2.2.1, page 23).

Best supportive care and HSCT were not considered as comparators for people with relapsed or refractory APL in the CS.

3.4 Outcomes

The NICE final scope lists the following outcome measures:

- overall survival (OS)
- progression free survival (PFS)
- response rates (bone marrow remission)
- adverse effects of treatment (AE)
- health-related quality of life (HRQoL)

These outcomes are reported in the CS with one exception: PFS; instead event-free survival (EFS) was used. In the APL0406 trial, EFS was assessed at two years after diagnosis, with treatment failure defined as any of the following: 1) no achievement of haematologic CR after induction; 2) no achievement of molecular CR after three consolidation courses; 3) molecular relapse; 4) haematological relapse, or 5) death. EFS is similar to PFS in this instance.

3.5 Other relevant factors

The company states that "Given the high rates of overall survival achieved with APL treatments, ATO is unlikely to meet the end-of-life criteria." (CS, Page 91). The ERG agrees, this STA does not meet the end-of-life criteria.

There is no Patient Access Scheme (PAS) application.

The company states that: "making ATO available on the NHS is likely to allow a greater number of elderly patients to be treated, which may be an important step towards addressing the topical issue of under-treatment among elderly oncology patients" (CS, B1.4, page 20). In addition, the company mentions Jehovah's Witness patients as a potential equality concern. No further equity or equality issues were mentioned in the CS.

4. CLINICAL EFFECTIVENESS

4.1 Critique of the methods of review(s)

The company conducted two systematic reviews to identify evidence on the clinical effectiveness and safety of ATO and other treatments for adults with APL. One review focused on randomised controlled trial (RCT) evidence and the other on non-RCT evidence. The non-RCT evidence was intended to inform the use of ATO as first line treatment only. This section critiques the methods of the reviews including searching, inclusion criteria, data extraction, quality assessment and evidence synthesis.

4.1.1 Searches

The company submission stated that in order to address the decision problem two separate searches were conducted in July 2016 which were then updated in October 2017. One search was designed specifically to identify RCTs, whilst a second search was conducted to identify non-RCTs "in order to provide the widest possible range of data."[1] Search strategies were reported in detail in Appendix D of the company submission for the following databases: MEDLINE, MEDLINE in-Process, Embase, and Cochrane Central Register of Controlled Trials (CENTRAL). The host provider was reported for MEDLINE and Embase, but not for CENTRAL. The date the searches were conducted was provided, though the date span of the databases searched was not. In response to the ERG clarification letter the company provided the database date of inception, and the date the searches were conducted, but not the date span. Searches utilised study design filters based on the Cochrane Highly Sensitive Search Strategy for identifying randomised trials (although this was not explicitly reported).[22] It is not clear where the study design filters were derived from for the non-RCTs searches. Searches of the trials register ClinicalTrials.gov were also conducted.

Additional searches of the following conference proceedings were reported in the main text of the company submission (section B.2.2.1) for 2011-2017: American Society for Clinical Oncology (ASCO), American Society of Hematology (ASH) and European Hematology Association (EHA). However, no details of the conference proceedings search strategies, date of searches or results were provided in Appendix D. Details of the conference proceedings searches were provided in response to the ERG clarification letter: search terms used, dates of the conferences searched, and number of abstracts retrieved.

ERG comment:

- Relevant studies could have been missed due to sub-optimal use of proximity operators, truncation and synonyms in search strategies. The eligibility criteria provided in Table 2.1 of the company submission included systematic reviews and meta-analyses, but no attempt to search for these study designs was made.
- The search strategy provided in Appendix D of the CS reported a simultaneous search of MEDLINE and Embase using the Ovid interface without including both MeSH and EMTREE subject headings. Search filters were used for the wrong databases and safety data may have been missed because study design terms used to search for non-RCTs were possibly too restrictive to capture all safety data.
- It is possible that potentially relevant studies were excluded from the final search results because the method used to limit the MEDLINE and Embase searches to human studies was incorrect. See Appendix 1 for further details.

4.1.2 Inclusion criteria

As stated above, the company conducted two systematic reviews to identify evidence on the clinical effectiveness and safety of ATO and other treatments for adults with APL. One review focused on RCT evidence and the other on non-RCT evidence. The non-RCT evidence was intended to inform the use of ATO as first line treatment only. The eligibility criteria used in the search strategy for RCTs and non-RCTs are presented in Table 4.1. The CS stated that two independent reviewers screened the studies identified through the searches, in order to determine the eligibility of each study. The two lists of selected references were then compared and all disagreements were solved by discussion, or if persistent, by a third reviewer.

Table 4.1: Eligibility criteria used in the review search strategy

	Review of RCTs	Review of non-RCTs
Population	*Inclusion Criteria* Adult participants with APL, aged ≥16 years, of both genders *Exclusion criteria* • Paediatric-only population • High-risk newly diagnosed APL • Significant cardiac comorbidities • Significant pulmonary comorbidities • Active non-APL malignancy • Pregnant women • Women who were breastfeeding during the time of the study	*Inclusion Criteria* Adult participants with APL, aged ≥16 years, of both genders *Exclusion criteria* Paediatric-only population (aged ≤ 15 years)
Interventions and Comparators	*Inclusion Criteria:* Any intervention *Exclusion criteria:* None	
Outcomes	*Inclusion Criteria* • Adverse events • OS • EFS • DFS or RFS • Cumulative incidence of relapse • Response rates (complete haematological and molecular remission rates) *Exclusion criteria:* None	
Study design	*Inclusion Criteria* RCTs, Phase II/III studies, systematic literature reviews of RCTs, or meta-analysis *Exclusion criteria* Opinion, editorial letter	*Inclusion Criteria* • Observational study • Cohort study • Prospective study (non-RCT) • Patient registry • Cross sectional study • Case-control study • Cases series including ≥ 6 cases *Exclusion criteria* • Opinion, editorial letter • RCTs • Case reports • Case series with ≤5 cases

	Review of RCTs	Review of non-RCTs
Other	*Inclusion Criteria:* None *Exclusion criteria* • Old conference abstracts: conference abstracts published prior to 2014 were excluded. • No full text available online. • Chinese articles published in non-core journals were excluded.	*Inclusion Criteria:* None *Exclusion criteria* • Old conference abstracts: conference abstracts published prior to 2014 were excluded • Studies including a population of <50 patients • Studies that did not include ATO in first line
Source: CS, Tables 2.1 and 2.2 DFS = EFS = event-free survival; OS = overall survival; RCT = randomised controlled trials		

ERG comment:

- Two reviewers were involved in the selection of studies for the reviews which helps to minimise bias.
- The ERG queried the exclusion of non-RCT "studies that did not include ATO in first line". The company provided a list of the 70 non-RCT studies excluded on this basis and stated they were "generally supportive of its use in this indication".[19] Further details of the company's response is provided in section 4.2.1. The ERG considers that non-RCTs could have been included for the relapsed/refractory population particularly as no directly relevant RCT evidence is presented (see section 4.2.1). The committee will need to consider whether it is necessary to explore the evidence further given the company's view that "the use of ATO in the relapsed or refractory APL setting is already so well-established in routine clinical practice that it would be difficult to provide NICE with novel information based on the analysis of additional studies."[19]
- The company further stated in the CS that "Chinese articles published in non-core journals were excluded, due to their frequently poor quality. Furthermore, Trisenox® is not marketed in China, so Chinese studies may be expected to report on the use of other ATO formulations. Nonetheless, relevant Chinese articles that met the inclusion criteria are summarised in Appendix L."[1] The ERG examined the Chinese RCTs that met the inclusion criteria and believes that the company emphasised the most relevant RCTs at first line in a UK setting. The Chinese trials used different treatment regimens when compared to APL0406, the main relevant trial in the CS. Therefore, it was reasonable to exclude them from more detailed analysis.

4.1.3 Critique of data extraction

The CS stated that one reviewer extracted relevant data from included studies and the results were reviewed by a senior manager for quality control.

ERG comment: Data extraction appears to have been conducted appropriately.

4.1.4 Quality assessment

The CS did not explicitly state that two reviewers were involved in assessment of trial quality. However, given that study selection and data extraction included two reviewers it is assumed that this process also included two reviewers to minimise risk of bias. Quality was assessed using a tool adapted from the Centre for Reviews and Dissemination's (CRD's) guidance for undertaking reviews in health care.[23] Elements assessed were randomisation, allocation concealment, comparability of groups, blinding of care providers, patients and outcome assessors and drop out, selective reporting of outcomes and use of intention to treat analysis and appropriate methods for dealing with missing data. The three main trials

(APL0406, AML17 and Raffoux et al.) were quality assessed using published papers as the company was not involved in the trials.

ERG comment: Study quality appears to have been assessed appropriately. Results of the company's quality assessment and the ERG's assessment of APL0406 are presented in section 4.2.4. We have not presented an assessment of AML17 as the intervention was not delivered according to the licence and therefore not of direct relevance to the decision problem. Neither have we assessed Raffoux et al. as this trial was not considered as meeting the NICE scope.

4.1.5 Evidence synthesis

The authors stated that as the trials included in the review used different comparators a network meta-analysis (NMA) would be most appropriate. However, after evaluation of the included studies, the authors concluded that an NMA was not feasible for any of the outcomes. Studies that were comparable in terms of time point and outcome had no mutual comparator for inclusion in a network.

ERG comment:

- The two trials identified for newly diagnosed patients had different dosing and regimens for the intervention arm and as only one of these trials was in accordance with the licence and therefore of direct relevance to the decision problem (APL0406) it would not be possible to conduct a meta-analysis in this population. Additionally, AML17 included 57 high risk patients, a population which is not part of the NICE scope, although subgroup analysis was conducted by risk.
- For patients with relapsed/refractory disease one trial only was identified, which was not relevant for the decision problem as ATO was included in both treatment arms; therefore, a meta-analysis could not be performed in this population either.

4.2 Critique of trials of the technology of interest, their analysis and interpretation (and any standard meta-analyses of these)

4.2.1 Overview of the evidence in the submission

Newly diagnosed patients

Two main trials were included in the submission for patients with newly diagnosed APL (APL0406 and AML17).[13, 14] Both of these compared AATO (ATRA+ATO) to AIDA. Both trials focused on adults (age 18 in APL0406 and age 16 in AML17). Both were RCTs and both were open label. APL0406 took place in Italy and Germany whereas AML17 had trial centres in the UK, Denmark and New Zealand. AML17 also included patients at high risk who do not form part of the scope of this submission. Both trials presented a final analysis of patient outcomes at 53 months. However primary outcomes differed. APL0406 assessed event-free survival (EFS) at two years after diagnosis whilst AML17 assessed quality of life outcomes. Both considered a range of secondary outcomes including overall survival.

Relapsed/refractory APL

The only trial presented in the CS relating to relapsed/refractory patients was Raffoux et al.[18] All patients had been previously treated with ATRA and anthracycline-based chemotherapy. The trial compared AATO with ATO alone which is not a relevant comparison according to the NICE scope. The trial had just 20 patients and a median follow up of 21 months. No trials compared ATO regimes with hematopoietic stem cell transplantation or with best supportive care as specified in the NICE scope.

No relevant comparative trials of ATO alone were presented for either newly diagnosed patients or those with relapsed/refractory disease.

An overview of the three main trials in the CS is presented in Table 4.2.

Table 4.2: Overview of RCTs in the submission

Trial name	APL0406	AML17	Raffoux et al (2003)
Population	Patients with newly-diagnosed, low to intermediate risk APL aged 18 to 71 years	Patients with newly-diagnosed APL, of any risk group aged ≥ 16 years	Patients with APL in first or subsequent relapse, aged ≥ 12 years. All previously treated with ATRA and anthracycline-based chemotherapy.
Intervention	AATO	AATO	AATO
Comparator	AIDA	AIDA	ATO alone
Outcomes	Primary: EFS at 2 years after diagnosis Secondary: • Rate of haematological CR after induction • Rate of molecular CR after 3 consolidation cycles • Probability of OS • Cumulative incidence of relapse • Toxic effects • QoL	Primary: QoL (EORTC QLQ-C30 and HADS) Secondary: • OS • RFS • EFS • Incidence of relapse (morphological and molecular)	Primary: 2 week reduction in time to haematological CR Secondary: • Safety • Molecular response • OS • DFS
Trial design and duration	Prospective, randomised, open-label, phase III non-inferiority trial	Randomised, controlled, phase III open-label trial	Randomised study
Median follow up	Initial cohort: 34.4 months (updated analysis 53 months) Final cohort: 40.6 months	30.5 months (53.4 months in updated analysis)	21 months
Location	40 centres in Italy and 27 in Germany	81 hospitals in the UK, Denmark and New Zealand	Details not reported. Patients were referred onto the study from 17 hospitals in France.
Number of participants	156 in initial cohort and 266 in final cohort	235 randomised patients	20

Source: Tables 2.3, 2.4 and 2.5 of the CS
AATO = ATRA+ATO; AIDA = ATRA + idarubicin; APL = Acute promyelocytic leukaemia; ATO = arsenic trioxide; ATRA = All-trans retinoic acid; CR = complete remission; DFS = disease-free survival; EFS = event-free survival; EORTC = European Organisation for Research and Treatment of Cancer; HADS = Hospital Anxiety and Depression Scale; OS = overall survival; QoL = quality of life; RCT = randomised controlled trial; RFS = relapse-free survival

The target sample size for the APL0406 trial was 162 patients at which point randomisation and enrolment were closed. This represented the initial cohort of patients. However, it was found, on preliminary analysis, that compliance with quality of life assessment was suboptimal. In order to ascertain the effects of arsenic-based treatment on quality of life the protocol was amended to increase the sample size to 276 patients (a final cohort). It is important to realise that the initial cohort of patients are included in the final cohort. Numbers available for analysis in the initial cohort were 156 and 266 in the final cohort.

Newly diagnosed patients taking part in AML17 did not receive ATO at its licensed indication. Additionally, gemtuzumab ozogamicin (GO) was an optional treatment in high-risk patients randomised to AATO and seven low-to-intermediate risk patients in this study received GO to counteract rising white blood cell (WBC) counts. Treatments given in AML17 are shown in Table 4.3 and for APL0406, which used ATO according to its licensed indication, in Table 4.4.

Table 4.3: Overview of treatments in AML17

Intervention	AATO (ATRA+ATO)	AIDA
Induction	Oral ATRA (45 mg/m^2/day until CR or for up to 60 days) + IV ATO (0.3 mg/kg on days 1–5 and 0.25 mg/kg twice-weekly in weeks 2–8) Gemtuzumab ozogamicin (6 mg/m^2 single IV infusion within days 1–4).[1]	Oral ATRA (45 mg/m^2/day until CR or up to 60 days) IV idarubicin (12 mg/m^2/day for a total of 4 doses)
Consolidation	Cycles 1–3: oral ATRA (45 mg/m^2/day for 15 days, two weeks on, two weeks off) + IV ATO (0.3 mg/kg on days 1–5 and 0.25 mg/kg twice-weekly in weeks 2–4) Cycle 4: oral ATRA (45 mg/m^2/day for 15 days) + IV ATO (0.3 mg/kg on days 1–5 and 0.25 mg/kg twice-weekly in weeks 2–4)	1st cycle: oral ATRA (45 mg/m^2/day for 15 days) + IV idarubicin (5 mg/m^2/day for a total of 4 doses) 2nd cycle: oral ATRA (45 mg/m^2/day for 15 days) + IV mitoxantrone (10 mg/m2/day for a total of 4 days) 3rd cycle: oral ATRA (45 mg/m^2/day for 15 days) + IV idarubicin (12 mg/m^2/day for 1 dose)
Maintenance	No maintenance phase	No maintenance phase

Source: Table 2.4 of the CS

1) Gemtuzumab ozogamicin (GO) was an optional treatment in high-risk patients randomised to AATO. Of 30 high-risk patients in this group, 28 (93%) received GO, with the remaining two patients given an anthracycline instead. Additionally, seven low- to intermediate-risk patients in this study received GO to counteract rising WBC counts.

AATO = ATRA+ATO; AIDA = ATRA + idarubicin; ATO = arsenic trioxide; ATRA = All-trans retinoic acid; CR = complete remission; IV = intravenous

Table 4.4: Overview of treatments in APL0406

Intervention	AATO (ATRA+ATO)	AIDA
Induction	Oral ATRA (45 mg/m^2/day) + IV ATO (0.15 mg/kg/day) Both continued until CR or up to 60 days	Oral ATRA (45 mg/m^2/day until CR or up to <60 days) + IV idarubicin (12 mg/m^2/day for a total of 4 doses)
Consolidation	Cycles 1–3: oral ATRA (45 mg/m^2/day for 15 days, two weeks on, two weeks off) + IV ATO (0.15 mg/kg/day 5 days per week, four weeks on, four weeks off) Cycle 4: oral ATRA (45 mg/m^2/day for 15 days) + IV ATO (0.15 mg/kg/day 5 days per week for four weeks)	1st cycle: oral ATRA (45 mg/m^2/day for 15 days) + IV idarubicin (5 mg/m^2/day for a total of 4 doses) 2nd cycle: oral ATRA (45 mg/m^2/day for 15 days) + IV mitoxantrone (10 mg/m^2/day for a total of 5 days) 3rd cycle: oral ATRA (45 mg/m^2/day for 15 days) + IV idarubicin (12 mg/m^2/day for 1 dose)
Maintenance	No maintenance	Oral ATRA (45 mg/m^2/day for 15 days every 3 months for 2 years, for a total of 6 courses) alternating with intramuscular or oral methotrexate (15 mg/m^2/week) + oral 6-MP (50 mg/m^2/day) for a total of 7 courses
Source: Table 2.3 of the CS AATO = ATRA+ATO; AIDA = ATRA + idarubicin; ATO = arsenic trioxide; ATRA = All-trans retinoic acid; CR = complete remission; IV = intravenous		

ERG comment:

Newly diagnosed patients

- The most important point to note is that only one directly relevant RCT is presented in the submission (APL0406). Newly diagnosed patients taking part in AML17 (a mainly UK-based trial) did not receive ATO at its licensed indication. For this reason, the remainder of this report focuses on APL0406 in newly diagnosed patients which was the main trial used in economic modelling. AML17 is briefly described under section 4.2.7 'Supporting evidence'.
- The population in APL0406 is relevant to the scope as it includes adults with low-to-intermediate risk APL.
- The intervention and comparator in APL0406 are relevant to the scope of this appraisal. ATO is delivered at its licensed indication.
- The outcomes in APL0406 included in the scope of this appraisal are assessed. Event-free survival is assessed rather than progression-free survival but these outcomes are similarly defined in APL0406.
- APL0406 is randomised, open-label, non-inferiority trial. The fact that the trial is open-label means that care providers, participants and outcome assessors are not blind to treatment allocation so in this respect bias can be introduced. A quality assessment of this trial is found in section 4.2.4.
- APL0406 follows an initial cohort of 156 patients up to a median of 53 months. The final cohort including all 266 patients is followed up to a median of 40.6 months.
- APL0406 is a multicentre trial with centres in Italy and Germany. There are no UK patients. The committee will need to consider the importance of this issue given that the treatment and comparator are relevant to the UK setting. The trial does not include a maintenance phase for

ATO + AIDA. The company clarified that "Primary market research commissioned by Teva in 2015 suggested that APL treatment in the UK does not include maintenance therapy."[19] Furthermore, the ERG notes that licensing for ATO does not specifically include a maintenance phase.[20]

- The evidence for the efficacy and safety of ATO + AIDA in patients with low-to-intermediate risk APL is based on 266 patients from the APL0406 trial.

Patients with relapsed/refractory disease

- The only trial presented in the CS relating to relapsed/refractory patients was Raffoux et al which included 20 patients (10 in each arm). The trial compared AATO with ATO alone which is not a relevant comparator according to the NICE scope (both arms include ATO). Therefore, no relevant evidence in patients with relapsed/refractory disease was presented in the CS for relapsed/refractory patients.
- In the clarification letter the company was invited to include all relevant non-RCTs of ATO if no RCTs were available for this patient group. The company had excluded non-RCT studies which did not address first line patients in the CS. In response the company stated "the available non-randomised second-line studies of ATO are generally supportive of its use in this indication"[19] and added "Among the studies on second-line ATO use that were initially identified by our literature search but later rejected as they did not focus on first-line indication, two deserve particular attention….".[19] They described a retrospective analysis of 25 patients with relapsed APL treated with ATO for remission induction [24] and a retrospective registry-based study from Japan showing that the annual number of autologous transplants among APL patients in second complete response (CR) increased approximately four-fold after ATO became commercially available in the country in late 2004; however, it was not clear how many patients in this study had actually used ATO.[25] It was unclear why these two particular studies had been chosen and whether other evidence supporting or refuting the use of ATO was available.
- In response to clarification the company stated "Overall, Teva feel that the use of ATO in the relapsed or refractory APL setting is already so well-established in routine clinical practice that it would be difficult to provide NICE with novel information based on the analysis of additional studies."[19] The committee will need to decide if this is acceptable particularly given the low numbers of patients expected to be treated at this stage. The company estimates that if ATO-based treatment were provided at first line the number of patients to be treated for relapsed disease would be approximately 10 to 16 patients in England.[1]
- No trials of ATO alone were presented for those with relapsed/refractory disease. The company stated that "We were unable to identify suitable efficacy data for ATO alone other than those published by Raffoux et al."[19] and added "Furthermore, according to all experts and especially to Dr Dillon (clinical expert for the UK), ATO alone is rarely used nowadays".[19] The committee will need to decide if they are in agreement with this perspective.
- It should also be noted that no trials in the CS compared ATO regimes with hematopoietic stem cell transplantation or with best supportive care as specified in the NICE scope.

4.2.2 Statistical analysis of APL0406

APL0406 was designed as a non-inferiority trial aiming to show that AATO was non-inferior to AIDA. This was interpreted as the experimental (AATO) arm being at most 5% inferior to the control (AIDA) arm in terms of the percentage of patients who were alive and failure-free at two years (EFS at two years).

Expected two-year EFS was 85% in the AIDA arm, based on the AIDA-2000 trial,[26] and 95% in the AATO arm, based on a previous non-randomised study.[27] The trialists calculated that 73 patients per treatment arm (146 in total) would be required based on a non-inferiority limit of 5%. This was increased to 162 to allow 10% loss to follow-up. The trial reached its target accrual in September 2010, at which point randomisation and enrolment were closed. However, based on a preliminary analysis of available quality of life data, the trial protocol was amended to increase the target accrual for the final cohort to 276 patients (57 additional patients per arm) to reach optimal quality of life (QoL) compliance.

Non-inferiority was assessed by estimating the two-sided 95% confidence interval for the between-group difference in crude rates of two-year EFS and was confirmed if the lower bound was ≥-5%. The trialists conducted a sensitivity analysis that addressed all relevant scenarios for the patients who could not be evaluated, assuming poor outcome for all patients, favourable outcome for all patients, or poor outcome for patients in the AATO group and favourable outcome for those in the AIDA group.

All efficacy analyses in the APL0406 trial were stated to be based on the 'intention-to-treat (ITT)'principle, comparing groups according to the randomly assigned treatment. This was defined as all patients who received at least one dose of assigned therapy following randomisation (n=156 in the initial cohort, n=266 in the final cohort). A per-protocol non-inferiority analysis was also carried out for the primary efficacy endpoint (EFS at two years). The per-protocol analysis set included 229 patients with sufficient follow up (>24 months).

EFS was assessed by comparing Kaplan–Meier curves, taking into account time to treatment failure and loss to follow-up. Survival distributions (EFS, OS and DFS) were estimated with the use of the Kaplan–Meier product-limit estimator and compared between groups using a log-rank test. Cumulative incidence of relapse was compared between groups using the non-parametric Gray K-sample test. Differences in percentages and other categorical variables (response rates, toxicity) were compared using Fisher's exact test or a chi-squared test. Continuous variables were compared using Mann-Whitney and Kruskal-Wallis tests. All tests were two-sided.

HRQoL was a secondary end point of the APL0406 trial. The European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire–Core 30 was used to assess HRQoL at end of induction and after consolidation therapy. All analyses were based on those 156 patients (the initial cohort) who received at least one dose of treatment, with groups defined according to randomly assigned treatment. Primary analysis was performed, estimating mean HRQoL score over time and differences between treatment arms using a linear mixed model.[28]

ERG comment:

- Although APL0406 was designed as a non-inferiority trial, trialists were able to demonstrate the superiority of AATO at least on certain outcomes.
- Analyses appeared to have been conducted appropriately. However, it should be noted that an ITT analysis should normally be conducted on all patients randomised to an intervention whether or not any treatment was received. In this case the analysis of the final cohort in regard to EFS was conducted for 263 of 266 randomised.
- According to the CS, the APL0406 trial protocol was amended to increase the target accrual for the final cohort to 276 patients (57 additional patients per arm) to reach optimal quality of life (QoL) compliance. However, all QoL analyses were based on the initial cohort of 156 patients who received at least one dose of treatment.

4.2.3 Participants in APL0406

Table 4.5 shows the inclusion and exclusion criteria for the APL0406 trial.

Table 4.5: Participant inclusion and exclusion criteria in APL0406

Inclusion criteria	Exclusion criteria
• Age 18–71 years • Newly-diagnosed APL • Low- to intermediate-risk APL (WBC count at diagnosis $\leq 10\times10^9$/L) • Genetic confirmation of diagnosis required after initial enrolment* • WHO performance status score ≤2 • Creatinine level ≤3.0 mg/dL (≤265 μmol/L) • Bilirubin level ≤3.0 mg/dL (≤51 μmol/L)	• Age <18 and ≥71 • WBC count at diagnosis $>10\times10^9$/L • Other active malignancy at time of study entry • Lack of diagnostic confirmation at genetic level • Significant arrhythmias, ECG abnormalities** or neuropathy • Cardiac contraindications for intensive chemotherapy (L-VEF <50%) • Uncontrolled, life-threatening infections • Severe uncontrolled pulmonary or cardiac disease • Pregnancy*** or breastfeeding • Concomitant severe psychiatric disorder • HIV positivity • Use of other investigational drugs at the time of enrolment or within 30 days before study entry

Source; Table 2.6 of the CS

*Confirmation of diagnosis at genetic level was required for patient eligibility. However, to avoid delay in treatment initiation, patients were randomised on the basis of morphologic diagnosis only, before the results of genetic tests were available. APL diagnosis was genetically confirmed by one or more of the following methods: 1) detection of the PML–RARA fusion gene by RT-PCR, 2) demonstration of the t(15;17) translocation by conventional karyotyping or FISH, 3) evidence of a microspeckled PML pattern by indirect immunofluorescence assay

** Including: 1) congenital long QT syndrome, 2) history or presence of significant ventricular or atrial tachyarrhythmia, 3) clinically significant resting bradycardia (<50 beats per minute), 4) QTc >450 ms on screening EKG, 5) Right bundle branch block plus left anterior hemiblock, bifascicular block

*** Women who were either pregnant or breast feeding, or of child-bearing potential were excluded, defined as all women physiologically capable of becoming pregnant, unless they meet one of the following definitions: amenorrhea; post-surgical bilateral oophorectomy with or without hysterectomy; using a highly effective method of birth control (defined as those which result in a failure rate less than 1% per year) when used consistently and correctly, such as implants, injectables, oral contraceptives, IUDs, sexual abstinence or vasectomized partner.

APL = acute promyelocytic leukaemia; ECG = electrocardiogram; HIV = human immunodeficiency virus; L-VEF = left-ventricular ejection fraction;WBC = white blood count; WHO = world health organisation

The APL0406 trial included 266 patients with genetically confirmed newly diagnosed, low-to-intermediate risk APL. Table 4.6 shows the characteristics of the patients in the APL0406 trial. These include the initial cohort of 156 patients as results were presented for this group in addition to the final cohort. Details of patient characteristics are limited as the company did not conduct the trial and relied on published information for these data.

Table 4.6: Patient characteristics in APL0406

	APL0406 initial cohort		APL0406 final cohort	
Treatment arm	AATO (n = 77)	AIDA (n = 79)	AATO (n = 129)	AIDA (n = 137)
Male gender; n (%)	40 (52)	36 (46)	60 (46.5)	70 (51.1)
Age, years; median (range)	44.6 (19.1 to 70.2)	46.6 (18.7 to 70.2)	46.6 (18.8 to 70.2)	46.6 (18.0 to 70.3)
WBC count, x 10^9/L; median (range)	1.49 (0.32 to 10.00)	1.60 (0.30 to 9.61)	1.4 (0.3 to 10.0)	1.5 (0.3 to 9.6)
Platelet count, x 10^9/L; median (range)	31 (3 to 224)	27 (3 to 236)	36.5 (3 to 224)	31.5 (3 to 236)
Low risk, n (%)	33 (43)	27 (34)	57 (45.2)	55 (41.3)
Intermediate risk, n (%)	44 (57)	52 (66)	69 (54.7)	78 (58.6)
High risk, n (%)	NA	NA	NA	NA
Source: Table 2.7 of the CS[1, 12, 14] AATO = ATRA+ATO; AIDA = ATRA + idarubicin; ATO = arsenic trioxide; ATRA = All-trans retinoic acid; NA = not applicable; WBC = white blood cell				

The median age of participants in APL0406 was 46.6 years in both arms of the trial with ages ranging from 18 to 70 years. Just under half of the participants in APL0406 are male. Approximately 42% had low risk disease with the remainder having intermediate risk.

ERG comment: The ERG asked if the company had access to the clinical study report (CSR) for APL0406 but the company stated that as it was an investigator-sponsored study it was impossible for Teva to obtain additional data including the CSR. From the information available and using the AML17 trial as a proxy for UK practice, the ERG concludes that the patients appear to reflect those seen in UK clinical practice.

4.2.4 Quality assessment of APL0406

Quality was assessed in the CS using a tool adapted from CRD's guidance for undertaking reviews in health care.[23] Elements assessed were randomisation, allocation concealment, comparability of groups, blinding of care providers, patients and outcome assessors and drop out, selective reporting of outcomes and use of intention to treat analysis and appropriate methods for dealing with missing data. The company assessed the APL0406 trial using published papers as they were not involved in the trials. The ERG has also assessed the trial using the published papers. Results are shown in Table 4.7.

Table 4.7: Quality assessment of APL0406

Quality dimension	CS evaluation[1]	ERG evaluation[1]	ERG comment
Was randomisation carried out appropriately?	Not clear	Not clear	No information although the protocol states that 'central randomisation' was to be used.
Was the concealment of treatment allocation adequate?	Not clear	Not clear	No information
Were the groups similar at the outset of the study in terms of prognostic factors?	Yes	Yes	
Were the care providers, participants and outcome assessors blind to treatment allocation?	No	No	This was an open label trial.
Were there any unexpected imbalances in drop-outs between groups?	No	No	
Is there any evidence to suggest that the authors measured more outcomes than they reported?	Not clear	Not clear	Hospitalisation days were listed in the protocol but these do not appear to have been reported.
Did the analysis include an ITT analysis? If so, was this appropriate and were appropriate methods used to account for missing data?	Not clear[2]	No	Analysis is best described as 'modified ITT' as patients were required to have received at least one dose of assigned therapy after randomisation.

Source: Table 2.12 of the CS

1.Based on Platzbecker et al 2017[14] and Lo Coco et al 2013[12]

2. The ITT population was described as including all patients who received at least one dose of assigned therapy after randomisation, i.e. 266 and 156 patients in the final and initial cohorts, respectively. However, the ITT analysis for the primary endpoint actually included 263 and 150 patients, respectively, and the available information is insufficient to conclude if this analysis was appropriate and if appropriate methods were used to account for missing data.

ITT = intention to treat

ERG comment:

- It was not possible for the ERG to fully assess the quality of the trial without access to the full CSR. We agree with the company that issues relating to randomisation, allocation concealment and assessment of outcomes are unclear based on published information**.**
- The fact that the trial is open-label means that care providers, participants and outcome assessors are not blind to treatment allocation so in this respect bias can be introduced.
- Analysis was not strictly based on intention-to-treat as only patients who had received at least one dose of assigned therapy after randomisation were included in the analysis.

4.2.5 Results of APL0406

The main results for APL0406 are given in Table 4.8. The primary endpoint (EFS at two years for the initial cohort) showed that more patients were event-free at two years with AATO (97%) compared to AIDA (86%) ($p < 0.001$ for non-inferiority; $p = 0.02$ for superiority). Based on the final cohort of 129 patients receiving AATO and 137 receiving the AIDA regimen, AATO was found to be superior to AIDA. Significantly more patients were event-free ($p <0.001$) at two years with AATO (98.3%) compared to AIDA (86.8%) and at 50 months (97.3% vs. 80.0%).

Based on the final cohort, overall survival was significantly better ($p = 0.007$) in the AATO group (99.2% vs. 94.8%, $p = 0.007$) at two years and at 50 months (99.2% vs. 92.6% (87.9 to 97.5).

There were no statistically significant differences in the proportions of patients with haematological complete response after induction (100% vs 97%, $p = 0.12$) or in molecular complete response rate after third consolidation cycle, (100% vs 98.3%, p = NR) in the final cohort.

Quality of life results from the APL0406 trial are available only for the initial patient cohort (156 patients) assessed at the end of induction and following the third consolidation course. However, no pre-treatment baseline assessment was performed. Of 150 patients eligible for HRQoL assessment at the end of induction, 115 returned HRQoL forms (77%). After the third consolidation cycle 119 of 142 eligible patients (84%) returned forms. Compliance rates did not differ significantly between the two treatment arms. Measured on the EORTC QLQ-C30 (version 3), a significant overall difference between treatment arms was only detected for fatigue ($p=0.022$). The company stated that comparison of scores at individual time points showed that AATO was associated with significantly lower fatigue severity after induction but not after the third consolidation course. A long-term QoL analysis in the final APL0406 patient cohort remains to be reported (see section 4.2.8).

Table 4.8: APL0406: key clinical efficacy results

Endpoint and time frame	Initial cohort			Final cohort		
	AATO (n = 77)	**AIDA (n = 79)**	**P value**	**AATO (n = 129)**	**AIDA (n = 137)**	**P value**
EFS at 2 years, % (95% CI)	97 (NR)	86 (NR)	< 0.001 for non-inferiority; 0.02 for superiority	98.3 (95.9 to 100)	86.8 (81.1 to 92.8)	< 0.001
EFS at 50 months, % (95% CI)	96 (92 to 100)	81 (73 to 91)	0.003	97.3 (94.3 to 100)	80.0 (72.9 to 88.0)	< 0.001
OS at 2 years, % (95% CI)	99 (96 to 100)	91 (85 to 97)	0.020	99.2 (97.7 to 100)	94.8 (91.1 to 98.6)	0.007
OS at 50 months, % (95% CI)	99 (96 to 100)	88 (81 to 96)	0.006	99.2 (97.7 to 100)	92.6 (87.9 to 97.5)	0.007
DFS at 2 years, % (95% CI)	97 (94 to 100)	90 (84 to 97)	0.110	98.3 (95.9 to 100)	89.4 (84.1 to 95.0)	< 0.001
DFS at 50 months, % (95% CI)	NA	NA	NA	97.3 (94.3 to 100)	82.6 (75.6 to 90.3)	< 0.001
Haematological CR rate after induction, %	100	95	0.120	100	97.0	0.120
Molecular CR rate after 3rd consolidation cycle, n (%)	75 (100)	70 (100)	NR	115 (100)	117 (98.3)	NR
CIR at 2 years, % (95% CI)	1 (0 to 4)	6 (0 to 11)	0.240	0.9 (0 to 2.7)	8.2 (3.3 to 13.2)	0.0013
CIR at 50 months, % (95% CI)				1.9 (0.0 to 4.5)	13.9 (7.1 to 20.6)	0.0013

Source: Table 2.13 of the CS

AATO = ATRA+ATO; AIDA = ATRA + idarubicin; ATO = arsenic trioxide; ATRA = All-trans retinoic acid; CIR = Cumulative incidence of relapse; CR = complete remission; DFS = disease-free survival; EFS = event-free survival; OS = overall survival

4.2.6 Safety results of APL0406

The CS noted that all adverse events (AEs), adverse drug reactions (ADRs), serious adverse events (SAEs) and serious unexpected adverse reactions were recorded during the treatment period in the APL0406 study. No long-term safety data were collected. The company stated that "All patients in the APL0406 study received differentiation syndrome prophylaxis with prednisone (0.5 mg/kg/day) from day 1 until the end of induction treatment…. At the earliest manifestations of suspected differentiation syndrome (e.g., unexplained respiratory distress) temporary discontinuation of ATRA and/or ATO treatment and prompt administration of dexamethasone was recommended."[1] Adverse events in the final patient cohort of APL0406 are given in Table 4.9.

Table 4.9: Adverse events in the final patient cohort of APL0406

Adverse event	Time frame	AATO (n = 129)	AIDA (n = 137)	P value
Induction-specific adverse events, n (%)				
Patients with moderate to severe differentiation syndrome	During induction	21 (17)	17 (13)	0.38
Leukocytosis*	During induction	56 (43)	NR	NR
Haematological adverse events				
Patients with grade 3–4 neutropenia lasting >15 days, n (%)	During induction	61 (35)	109 (64)	< 0.001
	1st consolidation cycle	8 (16)	40 (67)	< 0.001
	2nd consolidation cycle	7 (7)	90 (92)	< 0.001
	3rd consolidation cycle	5 (15)	28 (85)	< 0.001
Patients with grade 3–4 thrombocytopenia lasting >15 days, n (%)	During induction	74 (38)	120 (62)	< 0.001
	1st consolidation cycle	6 (26)	17 (74)	< 0.001
	2nd consolidation cycle	6 (7)	77	< 0.001
	3rd consolidation cycle	8 (23)	16 (76)	< 0.001
FUO and infection episodes, n (%)	During induction	30 (23)	75 (55)	< 0.001
	1st consolidation cycle	10 (8)	8 (6)	0.540
	2nd consolidation cycle	4 (3)	46 (38)	< 0.001
	3rd consolidation cycle	2 (1.6)	2 (1.7)	1.000
Non-haematological adverse events				
Patients with QTc prolongation**, n (%)	During induction	11 (8.5)	1 (0.7)	0.002
	1st consolidation cycle	3 (2)	0	0.110
	2nd consolidation cycle	3 (2)	0	0.110
	3rd consolidation cycle	2 (1.5)	0	0.230
Patients with grade 3–4 hepatic toxicity, n (%)	During induction	51 (40)	4 (3)	< 0.001
	1st consolidation cycle	5 (4)	1 (0.7)	0.110
	2nd consolidation cycle	1 (0.8)	0	0.490
	3rd consolidation cycle	0	0	NA
Patients with grade 3–4 gastrointestinal toxicity, n (%)	During induction	3 (2)	25 (18.2)	< 0.001
	1st consolidation cycle	0	1 (0.8)	1.000
	2nd consolidation cycle	0	6 (4.9)	0.03

Adverse event	Time frame	AATO (n = 129)	AIDA (n = 137)	P value
	3rd consolidation cycle	0	0	1.000
Patients with grade 3–4 cardiac function abnormalities, n (%)	During induction	0	5 (3.7)	0.060
	1st consolidation cycle	0	0	NA
	2nd consolidation cycle	0	0	NA
	3rd consolidation cycle	0	0	NA
Neurotoxicity (all grades), n (%)	During induction	1 (0.7)	0	0.480
	1st consolidation cycle	5 (4.2)	0	0.020
	2nd consolidation cycle	6 (5)	0	0.010
	3rd consolidation cycle	7 (5.9)	0	0.006
Hypercholesterolemia, n (%)	During induction	14 (10)	12 (8.7)	0.550
	1st consolidation cycle	19 (16)	12 (9.6)	0.130
	2nd consolidation cycle	19 (16)	12 (9.7)	0.140
	3rd consolidation cycle	16 (14)	11 (9.0)	0.270
Hypertriglyceridemia, n (%)	During induction	29 (22)	29 (22)	0.760
	1st consolidation cycle	22 (18.4)	19 (15.2)	0.490
	2nd consolidation cycle	17 (14.4)	10 (8)	0.120
	3rd consolidation cycle	16 (14)	13 (11)	0.500

Source: Table 2.18 of the CS
* Leukocytosis was defined as WBC count >10 × 10^9/L
** Defined as QTc increased to >450 msec in males and >460 msec in females
AATO = ATRA+ATO; AIDA = ATRA + idarubicin; ATO = arsenic trioxide; ATRA = All-trans retinoic acid; FUO = fever of unknown origin

From Table 4.9 it can be seen that there were no significant differences between groups in numbers of patients with moderate to severe differentiation syndrome in the induction phase. However, in the AATO group there was a high incidence (43%) of leukocytosis during induction.

In the AATO group patients experienced fewer haematological adverse events including fever and infection episodes and grade 3 to 4 neutropaenia and thrombocytopaenia lasting over 15 days.

In terms of non-haematological adverse events, AATO was more favourable than AIDA for grade 3-4 gastrointestinal toxicity. However, a greater number of patients experienced QTc prolongation with AATO. This was particularly the case in the induction phase (8.5% vs 0.7%). A greater number of patients experienced grade 3 to 4 hepatic toxicity, again particularly in the induction phase (40% vs. 3%). In almost all patients, this toxicity was reversible and manageable with temporary drug interruption and dose adjustments as per protocol recommendations.[14] There were no instances of neurotoxicity with AIDA but 19 events were noted with AATO. Rates of hypercholesterolemia and hypertriglyceridemia were similar across groups.

ERG comment:

- Safety information on the AATO combination at the licensed dose for the first line treatment of APL is currently limited to one trial in which 129 patients have been exposed. Furthermore, the EMA commented that, “due to the potential synergistic toxicity of ATRA and ATO (i.e. on

hepatotoxicity), no direct extrapolation of safety data observed with single-agent ATO is considered adequate".[20]

- Knowledge of long-term toxicity of AATO is very limited. It is drawn to the attention of the committee that the EMA has recommended that the company conduct a post-authorisation long term safety cohort study. This is designed to explore further the long-term safety of AATO in newly diagnosed low-to-intermediate risk APL patients in a real-world clinical practice setting.
- The ERG draws to the attention of the committee the increase in rates of hepatotoxicity particularly during the induction phase. The EMA noted that this might be due to a possible synergistic toxic effect of ATRA and ATO. However, they noted that the observed hepatic damage was reversible with suspension of ATO and/or ATRA, and that no additional safety measures beyond a warning on the SmPC were necessary.[20]
- Patients will need to be carefully informed of the particular risks of the treatment regimen chosen.
- The company was asked to clarify a statement from the CS. They stated that "The estimated overall cumulative exposure to Teva Group products containing ATO was approximately 13,855 patients, with an estimated 363 patients exposed to ATO in 6 clinical trials sponsored by Teva Group." The ERG asked which six trials were being described and whether full data could be provided if relevant to the current decision problem. The company responded that the estimated cumulative clinical trials exposure to ATO in six clinical trials sponsored by Cephalon, Inc. (CTI 1073, CTI 1058, CTI 1061, ATO202, CTI 1064, C18477/3059/AM/ USCA) and 5 clinical trials sponsored by Cell Therapeutics, Inc. (CTI1057, CTI1059, CTI1060, CTI1062, CTI1063) was approximately 363 patients. The company (Teva) stated that they were aware of the fact that the cumulative number of patients exposed to arsenic trioxide in all clinical trials prior to the acquisition by Teva Group may be higher since, due to historical reasons, Teva's access to much of the data regarding studies conducted with ATO was limited.

Serious adverse events (SAEs) occurring in APL0406 are displayed in Table 4.10. Overall, 95 SAEs were reported in 65 patients: 43 SAEs in the AATO group and 52 in the ATRA + chemotherapy group.

Table 4.10: Serious adverse events in APL0406

System organ class **Preferred term, n (%)**	**AATO** **(n = 129)**	**AIDA** **(n = 137)**
Blood and lymphatic system disorders	**0**	**10 (7.3)**
Febrile neutropaenia	0	8 (5.8)
Bone marrow failure	0	1 (0.7)
Neutropaenia	0	1 (0.7)
Cardiac disorders	**3 (2.3)**	**7 (5.1)**
Pericarditis	1 (0.8)	2 (1.5)
Acute myocardial infarction	1 (0.8)	1 (0.7)
Cardiac failure	0	1 (0.7)
Ejection fraction decreased	0	1 (0.7)
Myocardial ischaemia	0	1 (0.7)
Syncope	1 (0.8)	0
Tachyarrhythmia	0	1 (0.7)
Eye disorders	**1 (0.8)**	**0**

System organ class **Preferred term, n (%)**	**AATO** **(n = 129)**	**AIDA** **(n = 137)**
Diplopia	1 (0.8)	0
Gastrointestinal disorders	**1 (0.8)**	**5 (3.6)**
Anal haemorrhage	0	1 (0.7)
Diarrhoea	0	1 (0.7)
Dyspepsia	1 (0.8)	0
Emesis	0	1 (0.7)
Inguinal hernia	0	1 (0.7)
Pancreatitis acute	0	1 (0.7)
General disorders	**1 (0.8)**	**5 (3.6)**
Mucusal inflammation	0	2 (1.5)
Pyrexia	1 (0.8)	2 (1.5)
Fever in aplasia	0	1 (0.7)
Hepatic disorders	**4 (3.1)**	**0**
Hepatotoxicity	1 (0.8)	0
Hypertransaminasemia	1 (0.8)	0
Hepatic failure	1 (0.8)	0
Cholelithiasis	1 (0.8)	0
Injury, poisoning and procedural complications	**1 (0.8)**	**1 (0.7)**
Maternal exposures before pregnancy	1 (0.8)	1 (0.7)
Infections and infestations	**6 (4.7)**	**10 (7.3)**
Pneumonia	2 (1.6)	2 (1.5)
Bronchopneumonia	0	2 (1.5)
Catheter site infection	2 (1.6)	0
Infection	0	2 (1.5)
Sepsis	0	2 (1.5)
Febrile infection	1 (0.8)	0
Herpes zoster	1 (0.8)	0
Bacteraemia	0	1 (0.7)
Urinary tract infection	0	1 (0.7)
Investigations	**7 (5.4)**	**2 (1.5)**
Electrocardiogram QT prolonged	2 (1.6)	0
ALT increased	2 (1.6)	0
AST increased	1 (0.8)	0
Hepatic enzyme increased	1 (0.8)	0
C-reactive protein increased	1 (0.8)	0
Hyperglycaemia	0	1 (0.7)
Transaminases increased	0	1 (0.7)

System organ class Preferred term, n (%)	AATO (n = 129)	AIDA (n = 137)
Nervous system	**4 (3.1)**	**1 (0.7)**
Cerebrovascular accident	1 (0.8)	0
Cerebral haemorrhage	1 (0.8)	0
Depression	1 (0.8)	0
Hydrocephalus	1 (0.8)	0
Ischaemic stroke	0	1 (0.7)
Psychiatric disorders	**1 (0.8)**	**0**
Confusional state	1 (0.8)	0
Reproductive system and breast disorders	**1 (0.8)**	**0**
Endometriosis	1 (0.8)	0
Respiratory, thoracic and mediastinal disorders	**10 (7.8)**	**7 (5.1)**
Retinoic acid syndrome	1 (0.8)	3 (2.2)
Respiratory failure	2 (1.6)	2 (1.5)
APL differentiation syndrome	3 (2.3)	0
Dyspnoea	3 (2.3)	0
Acute respiratory distress syndrome	0	1 (0.7)
Pneumonia	1 (0.8)	0
Pulmonary embolism	0	1 (0.7)
Skin and subcutaneous tissue disorders	**1 (0.8)**	**0**
Leucocytoclastic vasculitis	1 (0.8)	0
Vascular disorders	**1 (0.8)**	**4 (2.9)**
Extradural haematoma	0	1 (0.7)
Intracranial aneurysm	1 (0.8)	0
Pulmonary embolism	0	1 (0.7)
Shock haemorrhagic	0	1 (0.7)
Thrombosis	0	1 (0.7)

Source: EMA assessment report[20]
AATO = ATRA+ATO; AIDA = ATRA + idarubicin; APL = Acute promyelocytic leukaemia; ATO = arsenic trioxide; ATRA = All-trans retinoic acid

ERG comment: The ERG asked if information was available on treatment-related deaths in the included trials. The company responded that treatment-related deaths were not specifically reported. However, they provided additional information based on the trial publications. They stated that in APL0406, whilst no induction deaths were observed in the AATO group, in the AIDA group four patients died during induction therapy – two from differentiation syndrome, one from ischaemic stroke and one from bronchopneumonial causes. The company stated that as differentiation syndrome is a common adverse effect of ATRA (and ATO) these two cases, could be considered related to ATRA administration. In terms of the death from ischaemic stroke, they stated that this was reported in the publication as an SAE with a fatal outcome, and was deemed related to treatment with both ATRA and idarubicin. They further stated that the relationship between the death from bronchopneumonia and

study treatment was difficult to evaluate based on the information available. The company also stated that across both treatment groups, six patients died in complete remission (CR). One patient in the AATO group died of bronchopneumonia caused by infection with the H1N1 virus, (reported as unrelated to treatment with either ATRA or ATO). The remaining five patients who died in CR were in the AIDA group: bronchopneumonia (two, both considered related to treatment), and one each from haemorrhagic shock (unrelated), pulmonary embolism (unrelated) and secondary myelodysplastic syndrome (MDS) (reported as treatment-related).[19]

4.2.7 Supporting evidence

As the AML17 trial was conducted largely in UK patients, it is useful to compare its characteristics with those of APL0406. One difference is that while the APL0406 trial enrolled patients aged 18 to 71, the AML17 trial was open to patients aged 16 or over, with no upper age limit. Median age was similar (age 47) as were the proportion of male patients (51%). 235 patients were included in the trial of whom 57 were at high risk and 178 were at low risk. Patients in the low risk category in AML17 had a similar WBC (up to 9.9)[13] to those in APL0406.

It has already been discussed that the intervention in AML17 was not according to the licensed dose. This trial had less frequent arsenic dosing and higher dosage compared to APL0406. Additionally, gemtuzumab ozogamicin (GO) was an optional treatment in high-risk patients randomised to AATO and seven low-to-intermediate risk patients in this study received GO to counteract rising WBC counts. In contrast to APL0406, no prophylaxis for differentiation syndrome was recommended in the AML17 trial.

Primary outcomes also differed. APL0406 assessed event-free survival (EFS) at two years after diagnosis whilst AML17 assessed quality of life. The AML17 trial was due to recruit 300 patients, allowing more than 80% power to detect a difference of 6 to 7 points (out of 100) on the global health scale of the EORTC QLQ-C30 questionnaire based on data the AML15 trial.[29] However, AML17 closed randomisation after recruiting 235 eligible patients as no further drug supply was available. As a result, the trial had 80% power to detect a difference of 7.5 points on the global health scale of the EORTC QLQ-C30 questionnaire. The Hospital Anxiety and Depression Scale (HADS) was also used to assess quality of life. Patients enrolled in the AML17 trial returned a total of 671 completed QoL forms (156 at baseline, 137 at three months, 139 at six months, 136 at 12 months and 103 at 24 months). The company reported that no statistically significant difference was detected in the primary outcome of global functioning (effect size = 2.17 (95% CI: 2.79 to 7.12)). The company reported that, based on the power calculation, the confidence intervals ruled out a minimally clinically important disadvantage of six points for AATO compared with AIDA. For other measures, including fatigue, which was significantly better with AATO than AIDA in the APL0406 trial, benefits of AATO were of modest size and results not statistically significant. Small but statistically significant benefits of AATO over AIDA were seen for cognitive functioning (effect size = 5.95 (95% CI: 0.26 to 11.63)) and role functioning (effect size = 6.74 (95% CI: 0.26 to 13.21)). The remainder of the results are given in Table 4.11.

Table 4.11: AML17 key clinical efficacy results

Endpoint and time frame	AATO (n = 116)	AIDA (n = 119*)	HR or OR (95% CI)	P value
Haematological CR, NR, n (%)	109 (94)	106 (89)	OR 0.54 (0.21 to 1.34)	0.180
Molecular CR, NR n (%)	106 (91)	105 (88)	OR 0.71 (0.31 to 1.65)	0.430
OS at 4 years, % (95% CI)	93 (86 to 96)	89 (81 to 93)	HR 0.60 (0.26 to 1.42)	0.250
Early mortality at 30 days, % (95% CI)	4 (2 to 10)	6 (3 to 12)	HR 0.72 (0.23 to 2.31)	0.560
Early mortality at 60 days, % (95% CI)	5 (2 to 11)	9 (5 to 16)	HR 0.55 (0.21 to 1.43)	0.220
EFS at 4 years, % (95% CI)	91 (84 to 95)	70 (56 to 80)	HR 0.35 (0.18 to 0.68)	0.002
Haematological RFS at 4 years, % (95% CI)	97 (90 to 99)	78 (63 to 88)	HR 0.24 (0.09 to 0.60)	0.004
Molecular RFS at 4 years, % (95% CI)	98 (91 to 99)	70 (62 to 83)	HR 0.17 (0.08 to 0.39)	< 0.001
Cumulative incidence of death in remission at 4 years, % (95% CI)	2 (1 to 9)	1 (0.2 to 8)	HR 1.72 (0.18 to 16.6)	0.640
Cumulative incidence of haematological relapse at 4 years, % (95% CI)	1 (0.1 to 7)	18 (10 to 34)	HR 0.16 (0.06 to 0.46)	< 0.001
Cumulative incidence of molecular relapse at 4 years, % (95% CI)	0	27 (18 to 45)	HR 0.12 (0.05 to 0.30)	< 0.001
Cumulative incidence of tMDS -AML at 4 years, % (95% CI)	0	3 (0.4 to 17)	HR 0.15 (0.003 to 7.48)	0.340

Source: Table 2.16 of the CS
*No data were available for survival or relapse for two patients in the ATRA + chemotherapy group (1 low risk, 1 high risk)
AATO = ATRA+ATO; AIDA = ATRA + idarubicin; ATO = arsenic trioxide; ATRA = All-trans retinoic acid; CI = confidence interval; EFS = event-free survival; HR = hazard ratio; NR = not reported; OR = odds ratio; OS = overall survival; RFS = recurrence-free survival; tMDS-AML = treatment-related acute myeloid leukaemia or myelodysplastic syndrome

As for APL0406, EFS was superior in the AATO group (HR 0.35 (0.18 to 0.68). Both haematological and molecular RFS were superior in the AATO group. However, outcomes relating to early mortality and overall survival were not significantly different between treatment groups. This was in contrast to APL0406 where overall survival was superior for AATO.

ERG comment:

- The AML17 trial provides supporting evidence only for this submission as the intervention in AML17 was not according to the licensed dose.
- The patients in AML17 are predominantly from the UK so represent a population relevant to clinical practice. However, the trial includes high risk patients who are not relevant to this submission.

- A comparison between the results of APL0406 and AML17 is difficult due to differences in population, intervention and other factors such as provision of prophylaxis for differentiation syndrome in APL0406.
- AML17 provides additional evidence of the efficacy of the AATO regime for selected clinical outcomes.
- It is noted that the primary outcome of quality of life was not found to be superior for AATO but it is possible that the trial was underpowered to investigate this.

4.2.8 Ongoing trials

The company mentioned in the CS that a post-authorisation safety study (PASS) is in the process of approval by the EMA. This study will start in 2018 and run for five years to evaluate long-term safety in APL patients treated at first line with AATO.

The ERG asked if any further analyses were planned or publications in process regarding the trials in the CS and if any details were available on the quality of life assessment of the final cohort of APL0406. The company responded that 'APL0406 was an Investigator Sponsored Study and Teva only received the final publication. However, according to Professor Lo-Coco, a publication presenting the updated outcome of patients enrolled in the APL0406 trial at a 60-month median follow-up is planned for 2018.'[19] and "the long-term quality of life analysis will be based on a decision by the principal investigators, Prof. Efficace and Prof. Lo Coco. Teva is expecting the final publication of this analysis in 2019."[19]

ERG comment:

- The ERG is satisfied that none of the ongoing trials could have been used to inform the submission.
- The ERG notes that efficacy and safety of AATO for the treatment of patients at first line (beyond 50 months assessed in the trial) is unknown.
- Ongoing research will highlight longer-term efficacy, safety and quality of life issues.

4.3 Critique of trials identified and included in the indirect comparison and/or multiple treatment comparison

Not applicable.

4.4 Critique of the indirect comparison and/or multiple treatment comparison

Not applicable.

4.5 Additional work on clinical effectiveness undertaken by the ERG

No further work on clinical effectiveness was undertaken by the ERG.

4.6 Conclusions of the clinical effectiveness section

The CS included systematic reviews of the evidence for arsenic trioxide and its comparators in newly diagnosed and relapsed/refractory patients as per the NICE scope. The company presented evidence from three RCTs: Two of these were trials in newly diagnosed APL (APL0406 and AML17) and the third was a study in patients with relapsed APL (Raffoux, et al. 2003).

Untreated APL

Both trials in newly diagnosed APL (APL0406 and AML17) compared AATO (all-trans retinoic acid (ATRA) + ATO) with AIDA (ATRA + idarubicin). APL0406 included 266 patients with newly-diagnosed, low-to-intermediate risk APL aged 18 to 71 years; while AML17 included 235 patients with newly-diagnosed APL of any risk group, aged 16 or over (no upper age limit). APL0406 took place in Italy and Germany whereas AML17 had trial centres in the UK, Denmark and New Zealand. The dosing and regimens for the intervention arm (AATO) in AML17 were not in accordance with the licence; while the dosing and regimens for the intervention arm (AATO) in APL0406 were in accordance with the licence. There were further differences between populations (inclusion of high risk patients in AML17) and outcomes. It is not, therefore, appropriate to pool the results. In this report we focused primarily on APL0406 as this was according to the licence. The trial was not conducted by TEVA so both the company and the ERG relied on published information for details. This meant that specific issues regarding trial quality were not always clear to the ERG. There are no UK patients in APL0406. The committee will need to consider the importance of this given that the treatment and comparator are relevant to the UK setting. The trial does not include a maintenance phase.

Efficacy results from APL0406 showed that AATO significantly improved overall survival (OS) at 50 months compared with AIDA (99.2% vs 92.6% respectively, p=0.007). The primary endpoint of this trial was event-free survival (EFS) at two years in the initial cohort of 156 patients (97% with AATO vs 86% with AIDA, p<0.001 for non-inferiority, p=0.02 for superiority). EFS was significantly better in the AATO group across all subsequent analyses to reach 97.3% at 50 months in the full cohort of 266 patients, compared with 80.0% in the AIDA group (p<0.001). The primary source of the observed EFS benefit was a reduction in the number of relapses with AATO – at 50 months, the cumulative incidence of relapse was 1.9% in the AATO group compared with 13.9% in the AIDA group (p=0.0013). Efficacy results from AML17 were generally supportive.

Safety information on the AATO combination at the licensed dose for the first line treatment of APL is limited to 129 patients exposed to AATO in APL0406. In this trial in the induction phase there were no significant differences between groups in numbers of patients with moderate to severe differentiation syndrome but in the AATO group there was a high incidence (43%) of leukocytosis. In the AATO group patients experienced fewer haematological adverse events including fever and infection episodes and grade 3 to 4 neutropaenia and thrombocytopaenia lasting over 15 days. AATO was also more favourable than AIDA for grade 3-4 gastrointestinal toxicity. However, a greater number of patients experienced QTc prolongation with AATO particularly during induction (11% vs 0.7%). A greater number of patients experienced grade 3 to 4 hepatic toxicity, again particularly in induction (40% vs. 3%). In almost all patients, this toxicity was reversible and manageable with temporary drug interruption and dose adjustments as per protocol recommendations.[14] There were no instances of neurotoxicity with AIDA but 19 instances were noted with AATO. Patients will need to be carefully selected and informed of the particular risks of the chosen regimen. Knowledge of long-term toxicity of AATO for this group of patients is limited. It is drawn to the attention of the committee that the EMA has recommended a post-authorisation long term safety cohort study to explore this.

Relapsed or refractory APL

The CS presented one study in relapsed/refractory patients. The study by Raffoux et al. (2003) compared AATO with ATO, which is not a relevant comparison according to the NICE scope. OS was similar between the AATO and ATO study arms. Across both groups, the estimated two-year OS was 59% (95% CI: 35%–77%). EFS was not reported in this study.

The ERG considers that non-RCTs could have been included in the CS for the relapsed/refractory population particularly as no directly relevant RCT evidence is presented. The committee will need to consider whether it is necessary to explore the evidence further given the company's view that "the use of ATO in the relapsed or refractory APL setting is already so well-established in routine clinical practice that it would be difficult to provide NICE with novel information based on the analysis of additional studies."[19]

No trials of ATO alone were presented for those with relapsed/refractory disease. The committee will need to decide if they are in agreement with the company that ATO alone is rarely used in UK practice. It should also be noted that no trials in the CS compared ATO regimes with hematopoietic stem cell transplantation or with best supportive care as specified in the NICE scope.

5. COST EFFECTIVENESS

5.1 ERG comment on company's review of cost effectiveness evidence

Three SLRs were performed with the objectives to identify and select relevant 1) cost effectiveness analysis (CEA) studies in APL (CS Appendix G); 2) utility studies identify in APL (CS Appendix H); 3) costs and healthcare resource use studies in APL (CS Appendix I).

5.1.1 Searches performed for cost effectiveness section

The following paragraphs contain summaries and critiques of all searches related to cost effectiveness presented in the company submission.

Searches for cost effectiveness analysis review

A SLR was conducted to identify cost effectiveness evaluations. No details of the search methods used were provided in the main company submission (section B.3). Full details of the search strategies were reported in Appendix G for MEDLINE, Embase and the NHS Economic Evaluation Database (NHS EED). The host provider for MEDLINE and Embase was reported, but not for NHS EED. The company response to the ERG clarification letter confirmed that the Centre for Reviews and Dissemination (CRD) interface was used to search NHS EED. The date searches were conducted was provided, but not the database date range searched. Initial searches were conducted in July 2016, and an update search was conducted in October 2017.

Measurement and valuation of health effects

A SLR was conducted to identify health-related quality of life studies. No details of the search methods used were provided in the main company submission, section B 3.4.7. Full details of the search strategies were reported in Appendix H, although this was not indicated in the main company submission. MEDLINE, Embase and the NHS EED were searched. The host provider for MEDLINE and Embase was reported, but the host provider used to search NHS EED was not reported. The company response to the ERG clarification letter confirmed that the CRD interface was used to search NHS EED. The date searches were conducted was provided, but not the database date range searched. Initial searches were conducted in July 2016, and an update search was conducted in October 2017.

Cost and healthcare resource identification, measurement and valuation

A SLR was conducted to identify costs and resource use data for England. Details of the search methods used were not provided in the main company submission, section B 5. Full details of the search strategies were reported in Appendix I for MEDLINE, Embase and NHS EED. Searches were conducted in July 2016, and an update search was conducted in October 2017. The company submission reported that targeted searches were conducted to identify adverse event costs (per occurrence) if the required data were not available in the National Schedule of Reference Costs, 2014-15: these targeted searches were not provided. The company described how these data were identified via targeted searches in their response to the ERG clarification letter.

ERG comment: As per the clinical effectiveness search comments above (4.1.1), better use of adjacency, truncation and synonyms would have increased the sensitivity of the searches. Studies may have been missed due to inappropriate use of subject headings and search filters. Additionally, it is possible that potentially relevant studies were excluded from the final search results because the method used to limit the MEDLINE and Embase searches to human studies was incorrect. See Appendix 1 for further details.

5.1.2 Inclusion/exclusion criteria used in the study selection

Screening of publications by title and abstract was performed; followed by full publication review. Eligibility criteria for the review are presented in Table 5.1.

Table 5.1: Eligibility criteria for the systematic literature reviews

Eligibility domain	Inclusion criteria	Exclusion criteria
Population	Adult APL population	Children-only population (≤15 years)
Intervention(s)	Any intervention	-
Comparator(s)[a]	Any intervention	
Outcomes(s) 1 (Published economic evaluations)	Model structure (health states & transitions, decision tree), model specifications	
Outcomes(s) 2 (Utility studies)	Any relevant health utility data	
Outcomes(s) 3 (Cost/resource use studies)	Any relevant cost and resource use information	
Study design 1 (Cost effectiveness analysis studies)	Health economic evaluation, any methodology	Opinion, editorial letter
Study design 2 (Utility studies)	Any kind of study including utility data (utility elicitation studies or models referring to utility data)	
Study design 3 (Cost/resource use studies)	Any study including models, analysis of insurance databases or medical records, cross-sectional surveys, chart reviews or prospective observational studies	

ERG comment: The in- and exclusion criteria presented in Table 5.1 seem appropriate for the objective of this review. However, after considering the PRISMA charts, it appeared that additional exclusion criteria were applied. This included "Full text not available" and "Old study (>2 years)". As a result, some relevant studies might have been missed. Additionally, given the company eventually informed the model partly based on primary sources focusing on other populations than APL, extending the population for the SLR (beyond the APL population only) might have been informative.

5.1.3 Included/excluded studies in the cost effectiveness review

The searches related to CEA, utility and cost studies resulted in 145, 273 and 280 hits respectively (after removing duplicates) for screening. Eventually this resulted in six included publications for the review of CEA (of which two were abstracts), two for the utility review and 11 (of which four were abstracts) included publications for the review of cost studies respectively.

ERG comment: It is noticeable that publications were excluded based on "Full text not available" and "Old study (>2 years)", this was applicable to four, one and nine studies for the CEA, utility and cost studies SLRs respectively. As stated above, some relevant studies might therefore have been missed.

5.1.4 Conclusions of the cost effectiveness review

The cost effectiveness searches in the company submission were all documented and reproducible. However, there were a number of inconsistencies and inaccuracies, and some redundancy. The MEDLINE and Embase search strategies used an inappropriate 'animals' limit, and it is possible that relevant evidence may have been missed as a consequence

Considering the CEA SLR, the company concluded that in general, all of the included studies considered the cost effectiveness of AATO or ATO alone, compared to the combination of ATRA and chemotherapy. In all cases, the number of QALYs was higher in the groups receiving ATO than in the comparator groups. Mean total costs of AATO or ATO alone were higher than the costs of comparator treatments. Incremental cost effectiveness ratios (ICERs) differed between studies, which could be related to a number of methodological factors, including the fact that the studies concerned different countries.

Considering the utility SLR, the company stated that both included studies (which were also included in the CEA SLR) presented utilities which were based on conditions other than APL (i.e. chronic lymphocytic leukaemia and AML). Hence it was concluded that no utility values that were specific to APL could be identified from the SLR.

Considering the cost and resource use SLR, the company did not use the 11 identified studies in the economic model. This was justified by stating that the information captured was not compatible with that needed to populate the model, and in others by the fact NHS reference costs were preferentially used to ensure relevance to the current situation in England.

5.2 Summary and critique of company's submitted economic evaluation by the ERG

Table 5.2: Summary of the company's economic evaluation (with signposts to CS)

	Approach	Source/Justification	Signpost (location in CS)
Model	A Markov cohort model with 14 health states	To fully capture the impact of making ATO available to UK APL patients treated within the NHS	Chapter B 3.2
States and events	Health states include: - First line treatment induction, consolidation, < 2yrs and >2yrs remission health states - Second line treatment induction, consolidation and remission health states - HSCT (allogeneic or autologous) and post-HSCT remission and End of Life health states - tMDS/AML - Death	To capture the course of the disease, based on expert opinion.	Chapter B 3.2.2

	Approach	Source/Justification	Signpost (location in CS)
Comparators	AIDA	AIDA was the only comparator in the pivotal APL0406 trial, and the primary comparator in the AML17 trial. BSC was not used as a comparator because it only applies to the second line setting.	Chapter B 3.2.3
Population	Adult patients with newly diagnosed low-to-intermediate risk APL. Cost effectiveness was not assessed in the refractory / relapsed patients.	It is expected that clinical practice will shift towards the use of ATO as standard of care in newly-diagnosed patients.	Chapter B 3.2.1
Treatment effectiveness	Different sources are used to inform the different treatment effectiveness parameters. First line treatment effectiveness estimates (including probabilities of remission at different time points and probabilities of relapse) are derived from the APL0406 trial.[30] Second line treatment effectiveness estimates are derived from Raffoux et al (2003)[18] for probabilities of remission and treatment failure, and Tallman (2015)[31] for probabilities of relapse; and Russel et al (2017)[32] and Platzbecker et al (2016)[14] for probabilities of alloHSCT, autoHSCT. Post HSCT transitions are informed by Hosing et al (2003),[33] Ramadan et al (2012)[34] and de Botton et al (2005)[35] for mortality risk and by Holter Chakrabarty et al (2013)[36] for probability of molecular remission. The probability of death in the tMDS/AML state is informed by Ma et al (2007).[37] In addition to the RCT by Raffoux et al.,[18] the efficacy data in the second line part of the model were informed by clinical expert opinions and a previous cost effectiveness	The outcomes related to the use of ATRA+ATO and AIDA in newly-diagnosed APL were mainly estimated based on the head-to-head APL0406 clinical trial. A scenario analysis was also conducted with the treatment schedule and outcomes from the AML17 clinical trial. However, no head-to-head data versus AIDA were available for second line treatment.	Chapter B 3.3

	Approach	Source/Justification	Signpost (location in CS)
	model developed for ATRA+ATO in the US.[31]		
Adverse events	Several treatment-induced adverse events were considered in the model in terms of costs and patients QoL. Some could prompt treatment switch or discontinuation.	No justification for the selection of AEs was provided.	Chapter B 3.4.8
Health related QoL	Utilities are based on Tallman et al (2015),[31] which provided utility values in a chronic lymphocytic leukaemia population, and based on Lachaine et al (2015),[38] which reported utilities from acute myeloid leukaemia patients.	No studies reporting utility values specific to APL were identified. Previous cost-effectiveness studies in APL used proxy utilities for other conditions that the authors considered to be associated with utilities analogous to APL.	Chapter B 3.4.9
Resource utilisation and costs	Resource use and costs accounted for in the model are treatment acquisition costs, medical costs (treatment administration, supportive care, monitoring and follow-up, HSCT, palliative care), and management of adverse events costs. These were informed using NHS reference costs, the BNF, the PSSRU and publications of relevant trials.	The information captured in the 11 studies identified through the SLR was not compatible with that needed to populate the model, or not relevant to the setting in England.	Chapter B 3.5
Discount rates	Discount of 3.5% for utilities and costs	As per NICE reference case	Table 3.13
Sub groups	Not applicable		
Sensitivity analysis	Both DSA and PSA were performed as well as scenario analyses	As per NICE reference case	Chapter B 3.8

Source: CS[1]

AE = adverse events; AIDA=chemotherapy combined with all-trans retinoic acid; AML=acute myeloid leukaemia; APL=acute promyelocytic leukaemia; ATO=arsenic trioxide; ATRA=all-trans retinoic acid; BNF=British National Formulary; BSC=best supportive care; CS=company submission; DSA=deterministic sensitivity analysis; HSCT=haematopoietic stem cell transplant; PSA = probabilistic sensitivity analysis; PSSRU=Personal Social Services Research Unit; SLR=systematic literature review; tMDS/AML=treatment-related myelodysplastic syndrome or acute myeloid leukaemia.

5.2.1 NICE reference case checklist (TABLE ONLY)

Table 5.3: Summary of the company's economic evaluation (with signposts to CS): NICE reference case checklist

Elements of the economic evaluation	Reference Case	Included in submission	Comment on whether *de novo* evaluation meets requirements of NICE reference case
Population	As per NICE scope[2]	Partly	Cost effectiveness is not assessed in the refractory/relapsed (second line) setting.
Comparator(s)	Therapies routinely used in the National Health Service (NHS), including technologies regarded as current best practice	Partly	For the second line setting, as per the NICE scope, BSC should be considered as a comparator.
Type of economic evaluation	Cost effectiveness analysis	Yes	
Perspective on costs	NHS and Personal Social Services (PSS)	Yes	
Perspective on outcomes	All health effects on individuals	Yes	
Time horizon	Sufficient to capture differences in costs and outcomes	Partly	Time horizon of 40 years, used in the base-case, does not capture all relevant costs and effects
Synthesis of evidence in outcomes	Systematic review	Yes	
Measure of health effects	Quality adjusted life years (QALYs)	Yes	
Source of data for measurement HRQoL	Described using a standardised and validated instrument	Yes	
Source of preference data for valuation of changes in HRQoL	Time-trade off or standard gamble	No	HRQoL data used in the model are from studies in AML and CLL patients. Utility values were derived from cost effectiveness publications, not from the original studies.
Discount rate	An annual rate of 3.5% on both costs and health effects	Yes	
Equity weighting	An additional QALY has the same weight regardless of the other characteristics of the individuals receiving the health benefit	Yes	

Elements of the economic evaluation	Reference Case	Included in submission	Comment on whether *de novo* evaluation meets requirements of NICE reference case
Sensitivity analysis	Probabilistic modelling	Yes	
Source: CS[1] AML=acute myeloid leukaemia; CLL=chronic lymphocytic leukaemia; HRQoL=Health-related quality of life; NHS=National Health Service; PSS=Personal Social Services; QALY=quality-adjusted life year			

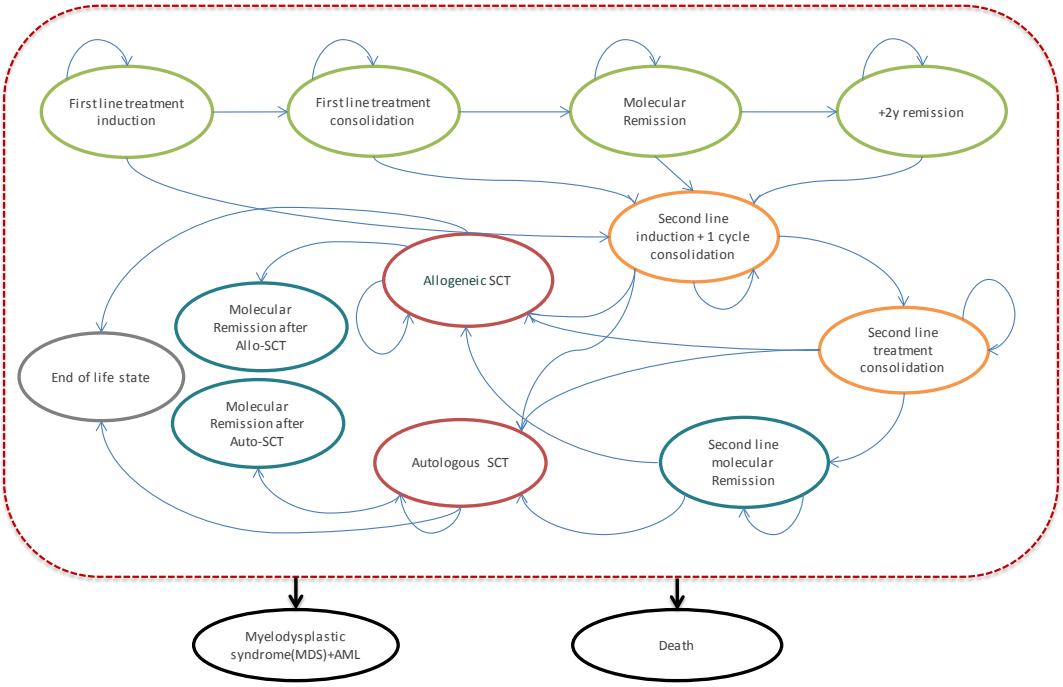
5.2.2 Model structure

The company developed a de novo Markov model comprising of 14 health states (or 69 when separately considering tunnel states, as indicated by the number of columns used in the Markov trace). No justification was provided for the chosen modelling approach, although the other economic evaluations identified in the SLR for CEA studies also used a Markov model structure. The number of health states deviated significantly from the four to five health states employed in the published economic evaluations, and this deviation was not justified.

The 14 health states of the model are shown in Figure 5.1. It should be noted that in Figure 5.1, curved arrows representing that patients can remain in health state, are missing for “End of life”, “Molecular remission after SCT”, “tMDS/AML”, “Death”. Furthermore, the arrow between “tMDS/AML” and “Death” is missing.

It is noteworthy that the length of time spent in some of these tunnel states depends on the treatment arm. Furthermore, the ATRA+ATO in first line treatment arm is implemented using two different subsequent treatment strategies: subsequent ATRA+ATO (AATO) in a large proportion of patients and subsequent AIDA in a small proportion of patients. Expert opinion suggested that the choice of subsequent treatment would depend on the duration of remission the patient has experienced. The first line AATO and second line AATO (subsequently referred to as AATO+AATO) strategy would be adopted in patients that had achieved two or more years of remission. The first line AATO, second line AIDA (AATO+AIDA) strategy is adopted in patients with less than two years remission. However, in the model, the proportion of patients experiencing two or more years of remission are implemented a priori, that is in different Markov traces.

Figure 5.1: Markov model structure with 14 health states



SCT= stem cell transplant; MDS+AML= myelodysplastic syndrome and acute myeloid leukaemia; y=year
Source: CS model file[1]

First line treatment health states

On treatment health states

There are two first line treatment health states in which patients are on treatment: the "first line induction" (during which patients are hospitalised) and the "first line consolidation" health states. All patients in the model are allocated to either AATO or AIDA first line treatment. For AATO "first line induction" consists of a maximum of two cycles of four weeks in the model and "first line consolidation" consists of a maximum of eight cycles of four weeks. In contrast, the maximum number of cycles in the "first line consolidation" state for the AIDA treatment arm is only three model cycles. As a note, the model implementation differs from the company's description in the CS, in which a maximum of three model cycles of induction and 10 cycles for consolidation phases was stated (although this might be technically possible in the economic model, first line treatment is restricted to fewer cycles).

From the "first line induction" state, patients can move to the "first line consolidation" state based on the median time necessary to achieve complete haematological remission before the maximum of two cycles and they can move to second line treatment if a cardiac event occurs. In contrast, patients would remain in the "consolidation" phase until the maximum of eight cycles, unless a cardiac event prompted treatment switch or they experienced tMDS/AML. At the end of the "first line consolidation" phase, patients that experienced treatment failure are moved to second line treatment.

Remission health states

There are two first line treatment remission health states: "first line molecular remission" and "+2y remission". In case of molecular remission after the "first line consolidation" phase, or, according to the company, if the patient could not be evaluated for remission (i.e. in the absence of evidence for treatment failure), the patient moves to the molecular remission health state. However, the latter is true

only for the AATO+AATO strategy in the model. There is an inconsistency in that, in the AIDA+AATO and AATO+AIDA strategies in the model, patients that could not be evaluated with PCR would be evaluated based on haematological response instead of being assumed to move to the molecular remission health state, and only in the case of haematological response would they move to the molecular remission health state.

Patients can remain in the "first line molecular remission" health state for a maximum of two years (24 model cycles, which is closer to 22 months due to the model cycle length of 28 days) and then move to the "+2y remission" health state. In the "first line molecular remission" health state, the probability of relapse is increased compared with the health state of "+2y remission". In case of a relapse in either one of these two remission health states, the patient moves to second line treatment. If there is no relapse, the patient remains in the "+2y remission" health state until death.

Second line treatment health states

Patients arrive in the second line induction phase in three cases: a) because they had experienced a cardiac event in first line induction or consolidation phases, b) because treatment failed after completion of the first line consolidation phase (40 weeks for AATO or 20 weeks for AIDA), or c) because of relapse when the patient had achieved molecular remission.

On treatment health states

There are two second line treatment health states in which patients are on treatment: the "second line induction + 1 cycle consolidation" and the "second line consolidation" health states. The "second line induction + 1 cycle of consolidation" health state consists of two model cycles of induction (mirroring the first line induction health state) and two model cycles of consolidation for AATO or one cycle of consolidation for AIDA in second line. Patients can move to the second part of this "second line induction + 1 cycle consolidation" health state (the consolidation cycle) if remission is achieved at one model cycle of induction therapy, to ensure that patients would always follow at least one cycle of consolidation. Patients in the AATO+AIDA strategy can transit from the "second line induction + 1 cycle consolidation" health state to the "tMDS/AML" health state. If, at the end of the "second line induction + 1 cycle consolidation" consolidation cycles, complete molecular remission is achieved, patients can continue consolidation treatment, or move to allogeneic or autologous HSCT. If complete molecular remission is not achieved, patients would undergo allogeneic HSCT (if they do not transit to "tMDS/AML"). Patients who experience a cardiac event discontinue second line treatment and undergo allogeneic HSCT.

The "second line treatment consolidation" health state comprises a maximum of six or two model cycles for AATO and for AIDA in second line respectively, at the end of which most patients undergo HSCT (allogeneic or autologous). Patients in the AATO+AIDA strategy can transit from the "second line treatment consolidation" health state to the "tMDS/AML" health state. Patients that cannot receive HSCT move to the "second line molecular remission" health state with no maintenance treatment.

Remission health state

There is only one second line remission health state: the "second line molecular remission" health state. Patients can stay in the second molecular remission health state until death, although there is a risk of relapse, which would prompt allogeneic HSCT. Patients can also move to allogeneic or autologous HSCT without experiencing a relapse.

Haematopoietic stem cell transplant related health states

The HSCT health states are populated with: a) patients that did not achieve molecular remission after the "second line induction + 1 cycle of consolidation" health state (only allogeneic HSCT), b) a proportion of patients that achieved complete molecular remission after the "second line induction + 1 cycle of consolidation" health state (autologous or allogeneic HSCT), c) patients that had a cardiac event in the "second line induction + 1 cycle of consolidation" or the "second line treatment consolidation" health states (only allogeneic HSCT), d) patients experiencing a relapse after achieving second molecular remission (only allogeneic HSCT), and e) proportions of patients in the "second line molecular remission" health states that did not experience a relapse (autologous or allogeneic HSCT).

Haematopoietic stem cell transplant health states

There are two HSCT health states: "alloHSCT" and "autoHSCT". The "alloHSCT" health state consists of six model cycles and reflects patients' hospitalisation and monitoring. Patients in this health state are at increased risk of acute graft versus host disease (GvHD) and at an increased risk of mortality compared with the general population. The "autoHSCT" health state consists of three model cycles and is also associated with an increased risk of mortality compared with that of the general population. Patients from both "alloHSCT" and "autoHSCT" health states would move to the respective remission after HSCT health states if the transplant was successful, or to the "End of life" health state if the transplant was not successful at the end of the respective tunnels.

Post HSCT health states

There are three post HSCT health states: "molecular remission after alloHSCT", "molecular remission after autoHSCT", and "End of life" (also called "Failure" in the model file). The "molecular remission after alloHSCT" is associated with increased costs, lower health-related quality of life, an increased mortality risk (compared with the general population) and a risk of developing chronic GvHD. The "molecular remission after autoHSCT" is associated an increased risk of mortality (compared with the general population), but with lower costs and better quality of life compared with the "molecular remission after alloHSCT", and no risk of chronic GvHD was applied. The "End of life" state is associated with low quality of life, high costs caused by extensive palliative care and a higher mortality risk than the molecular remission after HSCT health states.

Treatment-related MDS/AML and death health states

There are two other health states in the model: "tMDS/AML" and "death". Patients treated with AIDA in first or second line can experience tMDS/AML during the first line treatment consolidation phases. Patients stay there until they die, and mortality risk is increased compared to that of the general population.

Patients can die at any time in the model due to general population background mortality. Patients have an additional mortality risk when they are in the "tMDS/AML", the "alloHSCT" and "autoHSCT" health states, the "End of Life", the "molecular remission after alloHSCT" and the "molecular remission after autoHSCT" health states. The increased risk of mortality during induction and consolidation phases (due to bleeding and infection) was not modelled. Patients in the model therefore do not experience increased mortality in first and second line treatment induction, consolidation or remission health states.

ERG comment: The ERG's concerns relate to (a) a model structure that diverges from existing economic models in this therapeutic area, (b) inconsistent modelling of patients that cannot be evaluated for molecular remission, (c) of adverse events it was assumed that only cardiac events could prompt

treatment switch, (d) an error in the number of tunnels used to represent the two year molecular remission health state, (e) the absence of disease-related mortality from on treatment health states and (f) the applicability of alloHSCT to the UK clinical setting.

(a) A model structure that diverges from the one used in other economic evaluations in this condition. In response to the clarification question B1, the company justified the more complex model structure by stating that the existing economic evaluations did "not adequately reflect the trajectory of APL patients".[19] According to the company, the aim in this economic evaluation was to "offer more granularity with treatment phases, molecular remission and HSCT" to better reflect the clinical trajectory of APL patients. The company also explored the potential impact of their adopted model structure compared with the simpler model structures by comparing model outcomes and found, with the caveat that the settings in the model were different and a straight comparison is therefore not possible, that the inclusion of HSCT in this model and differences in drug costs across the models may account for differences in estimated costs between the models. The ERG considers that the model structure is appropriate to reflect this condition and treatment pathway.

(b) There is an inconsistency in what happens in the model when patients could not be evaluated for molecular remission. Patients in the AATO+AATO strategy would be assumed to be in molecular remission, while patients in the AATO+AIDA and AIDA+AATO strategies that could not be evaluated with PCR would be evaluated based on haematological response, and only if this was given patients were assumed to be in molecular remission (instead of assuming that all patients, regardless of haematological response, are in molecular remission). This was not justified and the ERG prefers to implement this in a consistent manner across treatment strategies. This is further explored in the treatment effectiveness section (Section 5.2.6) of this report.

(c) It was assumed in the model that among adverse events, only cardiac events could prompt a treatment switch. In response to the clarification question B3, the company stated that this was based on expert opinion. The company stated that "while it is possible that other serious AEs may prompt a treatment switch, they were not frequent enough to either find adequate probabilities or have any impact on the end results."[19] The minutes of the company's expert consultation, however, revealed that in a small proportion of patients reversible arrhythmia would also cause treatment switch. The ERG therefore has explored this in scenario analysis.

(d) The "first line molecular remission" health state is a tunnel state consisting of 24 model cycles. However, the company intended this to represent two years in remission. Due to the cycle length of four weeks, the appropriate number of cycles to reflect two years would be 26 cycles. In response to clarification question B19.c, the company acknowledged that these health states were missing and implemented 26 model cycles for this health state in a scenario, which resulted in slightly more favourable model outcomes for AATO, but with "minor impact".[19] The ERG implemented the 26 cycles in its base-case.

(e) No disease-related mortality was modelled during on treatment and remission phases. The company excluded disease-related mortality from the on treatment health states. In response to the clarification question B2, the company justified this modelling choice by stating that "the mortality rate observed during treatment in both the APL0406 trial and the AML17 trial was numerically lower for ATRA+ ATO compared to AIDA" and that the difference was not statistically significant.[19] The company explored the impact of adding disease-related mortality to the on treatment health states of their model and found that incremental QALYs increased and costs savings with AATO decreased. The ERG considers that the disease-related mortality risk is likely to be larger than the general population mortality risk in the treatment induction phase, which is an assumption consistent with the evidence

shown in the AML17 study. This additional mortality risk, is therefore implemented, during treatment induction, in exploratory analyses performed by the ERG.

(f) In the model, patients can undergo either autologous or allogeneic HSCT. However, it is questionable whether this is reflective of UK practice. For instance, patients who have a cardiac event in second line treatment can only receive allogeneic HSCT, not autologous HSCT. In response to clarification question B3.c, the company stated that this assumption was based on expert opinion that only patients with molecular remission are considered for autologous HSCT in the model and patients with a cardiac event would likely experience this before molecular remission. The company also explained that according to a UK expert "fewer HSCTs are conducted in the UK and allogeneic HSCT is generally not recommended in APL". The meeting minutes of the company's expert consultation support this. To reflect the uncertainty over the use of alloHSCT in the UK clinical practice, the ERG adopted a scenario in which only autoHSCT is performed.

5.2.3 Population

Arsenic trioxide (ATO), as per its marketing authorisation, is indicated for the treatment of:

- newly diagnosed low-to-intermediate risk APL (white blood cell count $\leq 10 \times 10^3/\mu l$) in combination with all-trans retinoic acid (ATRA) *(also referred to as first line treatment)*
- relapsed/refractory APL (previous treatment should have included a retinoid and chemotherapy) *(also referred to as second line treatment)*

The company only assesses the cost effectiveness of ATRA+ATO (AATO) in the newly diagnosed low-to-intermediate risk APL population, i.e. in first line treatment. The indication in the relapsed/refractory APL population is not assessed.

In the model, patients have an average age of 45 years, an average weight of 81kg and an average height of 169 cm. In total, 48.7% were assumed to be male. See Table 5.4 for the baseline characteristics of patients from the main evidence sources considered in the model.

ERG comment: AATO was only assessed in the newly diagnosed population (first line). In response to clarification question B5.a, the company provided an analysis in the relapsed/refractory population **(**second line) in which "the health states representing first line therapy were changed to second line, and those representing second line were neutralised (no transitions to these states were possible)".[19] The company further stated that "the analysis showed that ATRA+ATO was cost-effective versus AIDA in the second line setting with an incremental cost-effectiveness ratio (ICER) of £16,733 per QALY gained." The company's description of this analysis did not provide clarity over how this analysis was exactly performed. There was also a lack of clarity as to where the transition probabilities in the model were sourced from, and whether they reflected first line or second line treatment. The ERG therefore implemented their own scenario using the second line transition probabilities to reflect the relapsed/refractory population (second line).

Table 5.4: Key baseline patient characteristics in the APL0406 and AML17 trial

Study population	**APL0406 initial cohort**		**APL0406 final cohort**		**AML17**	
Treatment arm	**ATRA+ATO** (n=77)	**AIDA** (n=79)	**ATRA+ATO** (n=129)	**AIDA** (n=137)	**ATRA+ATO** (n=116)	**AIDA** (n=119)
Age, years; median (range)	44.6 (19.1–70.2)	46.6 (18.7–70.2)	46.6 (18.8–70.2)	46.6 (18.0–70.3)	47 (16–75)	47 (16–77)
Male gender; n (%)	50 (52%)	36 (46%)	60.0 (46.5%)	70.0 (51.1%)	60 (52%)	60 (50%)
WBC count, $\times 10^9$/L; median (range)	1.49 (0.32–10.00)	1.60 (0.30–9.61)	1.4 (0.3–10.0)	1.5 (0.3–9.6)	3.0 (0.4–100.9)	2.2 (0.4–78.2)
Platelet count, $\times 10^9$/L; median (range)	31 (3–224)	27 (3–236)	36.5 (3–224)	31.5 (3–236)	Not reported	Not reported
Low risk, n (%)	33 (43%)	27 (34%)	57.0 (45.2%)	55.0 (41.3%)	86 (74%)	92 (77%)
Intermediate risk, n (%)	44 (57%)	52 (66%)	69 (54.7%)	78 (58.6%)	Not reported	Not reported
High risk, n (%)	Not applicable	Not applicable	Not applicable	Not applicable	30 (26%)	27 (23%)

ATRA=All-trans retinoic acid; ATO=Arsenic trioxide; WBC=White blood cell

5.2.4 Interventions and comparators

First line therapy in APL generally consists of three consecutive treatment phases: induction, consolidation and maintenance.[1] However, maintenance treatment was not modelled and the justification provided by the company was that it is usually omitted in UK clinical practice with the aim of minimising the risk of treatment-related myelodysplastic syndrome or acute myeloid leukaemia (tMDS/AML) (CS Section B.1.3.2.1).

First and second line treatment with AATO was modelled with up to two cycles (of four weeks) of induction therapy followed by eight cycles (of four weeks) of consolidation therapy. Treatment protocols were in line with the APL0406 study.[12]

The only comparator used in the model was AIDA in first line. AIDA was implemented with up to two cycles (of four weeks) of induction therapy followed by three cycles (of four weeks) of consolidation therapy.

After first line treatment, subsequent treatment is prompted by relapse or by a cardiac event in the model. As per its marketing authorisation, ATO does not have to be administered in combination with ATRA in this second line population. However, the company only provided the analysis with ATRA+ATO in combination, in line with expert opinion stating that ATO alone would only rarely be used in the relapsed/refractory population.

Furthermore, there is a lack of clarity as to whether re-treatment with AATO can occur when a patient had relapsed. Informed by expert opinion, the company assumes in the economic model that patients who remained in remission for two years or longer following first line treatment with AATO would be re-treated with AATO upon relapse. Patients who achieved only a short (<2 years) remission after first line AATO treatment would be treated with AIDA. Of course, all patients treated with AIDA in first line, would switch to AATO in second line after relapse or cardiac event, independent of how long the period of remission was. In the model, the proportion of patients achieving two or more years of remission (and therefore assumed to be treated with AATO+AATO instead of with AATO+AIDA) are implemented a priori, that is in different Markov traces.

A comparison of ATO in the second line setting was not performed. If it had been implemented, according to the scope, Best Supportive Care should be considered a comparator in the second line setting. In response to the request for clarification, the company did perform an analysis of ATO at second line. However, Best Supportive Care was not implemented in the model as a comparator.

ERG comment: The ERG's concerns relate to (a) the lack of maintenance treatment in the company's model, (b) the absence of an analysis with ATO stand-alone in second line, and (c) the absence of BSC as a comparator in the second line setting.

(a) The company did not consider maintenance treatment in their model. In response to clarification question A.12, the company stated that their earlier insight based on market research, that in the UK maintenance therapy would only be provided in rare cases, was confirmed by a UK expert. It should be noted though that the rationale for maintenance therapy being rarely used in the UK is based on the AML15 study showing a higher incidence of tMDS/AML with maintenance therapy than the AML17. The incidences in both of these studies are for patients treated with chemotherapy regimens. Since the incidence of tMDS/AML is not a concern with AATO treatment, this does not justify not including maintenance treatment with AATO. However, maintenance therapy with AATO is not in line with the SmPC and therefore the ERG considers it as appropriate that maintenance therapy was not considered in the model.

(b) Only ATRA+ATO is modelled in second line, not ATO stand-alone. In response to clarification question B5.b, the company stated that there was no other evidence for second line treatment than the Raffoux et al study, which "did not show significant differences between ATO+ATRA and ATO alone, and, surprisingly, disease-free survival was better with ATO alone than with ATRA+ATO. Conducting this scenario would lead to better cost-effectiveness results for ATO vs. AIDA, reducing treatment acquisition costs without changing the effectiveness results."[19] The ERG was satisfied with this justification, especially given that experts stated that ATO alone would only rarely be used in UK clinical practice.

(c) BSC in second line was not included as a comparator in the model. In response to clarification question B5.c, the company stated that "all experts strongly stated that, due to the severity of the disease, best supportive care is not a relevant comparator in the second line setting, and that best supportive care is only a relevant alternative in 3rd or 4th line."[19] Furthermore, "given the very small number of affected patients, adding best supportive care as a comparator in 3rd or 4th line would have very little impact on the ICER."[19] The ERG was satisfied with this justification.

5.2.5 Perspective, time horizon and discounting

The model adopts the perspective of the NHS and Personal and Social Services (PSS) in England and Wales. The model time horizon is 40 years, at the end of which a significant proportion of patients in the model is still alive (> 40% of patients in the ATRA+ATO first line and AIDA second line arm). The model cycle length is 4 weeks to capture the treatment schedule and a half-cycle correction is applied. All costs and health gains were discounted at a rate of 3.5% per year.

ERG comment: The time horizon was too short to capture all relevant costs and outcomes. The company, in response to clarification question B20,[19] provided a scenario analysis with an extended time horizon of 56 years, which increased both cost savings and incremental QALYs. The ERG remains concerned about the long life expectancy of patients in the model and thinks that this calls the validity of the model into question (see section 5.2.12 for more details).

5.2.6 Treatment effectiveness and extrapolation

The treatment effectiveness section was structured according to different phases: first line, second line, haematopoietic stem cell transplant and other phases (i.e. tMDS/AML or death).

Treatment effectiveness of the AATO strategy was estimated by separately estimating Markov traces (as well as costs and QALYs) for AATO+AATO and AATO+AIDA. Subsequently, a weighted average was calculated with weights of 98% and 2% for AATO+AATO and AATO+AIDA respectively. No justification was provided for these weights.

First line health states

Transition from the first line health states were informed based on evidence from the APL0406 trial[12, 14] in combination with expert opinion (to inform assumptions related to the estimation of these transition probabilities).

From the "first line treatment induction" health state, patients can transit to "first line treatment consolidation" depending on the median time to CR (32 versus 35 days for AATO and AIDA respectively) and the occurrence of adverse events (i.e. cardiac event) requiring a treatment switch (probabilities of 0.0% and 3.0% for AATO and AIDA respectively). In case of adverse events requiring a treatment switch (i.e. cardiac events), patients transit to the "second line treatment induction phase +

1 cycle consolidation" health state. In case of no cardiac event after two cycles (i.e. 56 days), patients transit to "first line treatment consolidation".

From the "first line treatment consolidation" health state, patients can transit to the first line "molecular remission" health state if they have remission and do not transit to the "tMDS/AML" health state. For the Markov traces for AATO+AIDA and AIDA+AATO, the probability of remission is determined by the haematological response rate (98.4% versus 96.4% for AATO and AIDA respectively) for patients not evaluable with PCR (9.4% versus 9.8% for AATO and AIDA respectively). For patients that are evaluable with PCR, the above mentioned haematological response rate is multiplied by the molecular remission rate (100.0% versus 98.3% for AATO and AIDA respectively). Moreover, for the Markov trace for AATO+AATO, it was assumed that all surviving patients would be in remission.

Once the patients are in the first line remission health states (i.e. "molecular remission" and "+2y remission" health states), patients can transit to the "second line treatment induction phase + 1 cycle consolidation" health state based on the probability of relapse which was different for the first two years after remission and thereafter.

The transition probabilities from the first line health states (retrieved from the model) are presented in Table 5.5. This excludes general population mortality that is subsequently applied to the transition probabilities presented in this table (no additional mortality is assumed).

ERG comment: The ERG's concerns relate to (a) the overestimation of cardiac events and thus patients switching to second line induction for AIDA; (b) the calculation of patients transiting to first line consolidation early for AIDA; (c) calculations and assumptions regarding the remission probability; (d and e) assumptions and calculation concerning the relapse probabilities and; (f) not considering treatment switching due to reversible arrhythmia in the model. The transition probabilities that are adjusted in the ERG base-case are presented between square brackets in Table 5.5.

(a) The proportion of patients switching to second line induction due to experiencing a cardiac event during first line induction AIDA treatment is 4.6% in the economic model (based on the 3.0% probability per cycle) while only 3.7% (five out of 136) experience grade 3-4 cardiac events in the APL0406 trial. This overestimation (also reflected in Table 2.2 of CS Appendix J) was induced by the company using the median time to complete haematological remission (i.e. 35 days for AIDA) in order to convert this 3.7% to a cycle probability (of 3.0%). The ERG corrected this overestimation in its base-case by converting the 3.7% to a cycle probability of 2.4% using the average duration patients actually remained in the first line induction phase (in the model) for AIDA treatment (i.e. 44 days). This resulted in 3.7% of the patients switching to second line induction due to experiencing a cardiac event during first line induction AIDA treatment (consistent with the APL0406 trial).

(b) The proportion of patients transiting to first line consolidation early (i.e. after one cycle of induction) from first line induction AIDA treatment was calculated by using the median time to complete haematological remission of 35 days. However, for the calculation to convert this to a transition probability, the company assumed that the probability after 35 days (the median time) was 48.2% and not 50.0%. Presumably this was done to reflect patients switching treatment due to cardiac events. However, the ERG believes this is incorrect as the patients switching treatment due to cardiac events are already considered using separate transition probabilities. Hence this is corrected in the ERG base-case (implicitly assuming that patients switching treatment due to cardiac events have no 'early' response).

(c) In the CS it is assumed for AATO+AATO, that all surviving patients would transit to remission. This was done despite available evidence from the APL0406 trial to inform this parameter in the model. Given the lack of appropriate justification in the CS, the ERG preferred to inform the remission probability for AATO+AATO based on evidence from the APL0406 trial, consistent with what was done for AATO+AIDA and AIDA+AATO. Related to this, the probability of transiting to remission for patients that are evaluable with PCR was informed by the molecular remission rate in the ERG base-case for all strategies (removing the additional multiplication with the haematological response rate).

(d) The a priori division of first line AATO patients in the model into a group which will experience remission for ≥2 years (receiving AATO also in second line if necessary, AATO+AATO group) and another group that will not achieve this (receiving AIDA in second line, AATO+AIDA group) causes problems in the use of the probabilities of relapse in both parts of the model. For patients in the remission health states, the company assumed no relapse during the first 24 cycles for AATO+AATO, while for AATO+AIDA, the company assumed no relapse after 24 months. Although the company did not justify this approach, the ERG presumes that the company based this on the assumption that patients who remained in remission for ≥2 years following first line AATO would receive AATO+AATO, else patients would receive AATO+AIDA. However, irrespective of the exact rationale, the company should have used conditional probabilities of relapse when adopting this approach. The group of patients that are assumed to relapse during the first 24 months (AATO+AIDA) likely have a larger probability during this period than the average probability of all patients observed in the APL0406 trial (i.e. probability of relapse conditional on a relapse within two years after transiting to the remission health state). The company had intended that none of the patients should remain in remission for more than two years in this group of patients, which is not the case in the economic model. Similarly, for the group of patients that relapse after the first 24 months, the post 24 months relapse probability is likely higher than in the first 24 months (i.e. probability of relapse conditional on having no relapse within two years after transiting to the remission health state). By using unconditional relapse probabilities, the company likely underestimates the relapse probability for AATO. Given that the ERG did not have access to these conditional relapse probabilities, the average relapse probabilities were applied for AATO in the ERG base-case (i.e. 0.038% per cycle for the first two years and 0.042% per cycle thereafter) instead of the company's approach assuming 0% probabilities of relapse during and after the first 24 months for AATO+AATO and AATO+AIDA respectively.

(e) The 48 month relapse probabilities (used in the model to inform relapse more than two years after remission), were assumed to be equal to the 50 month relapse probabilities reported in the APL0406 trial publication. This was corrected in the ERG base-case by converting the 50 month probabilities to 48 month probabilities.

(f) The company submitted minutes considering the expert meeting the company organised to validate the CEA model. Although these minutes state that for AATO approximately 2% of patients experience reversible arrhythmia and therefore switch treatment, this was not incorporated in the model. Therefore, the ERG explored this scenario (using 2% cardiac events during the induction phase).

Table 5.5: Transition probabilities (per cycle) from the first line health states (excluding background mortality)

AATO+AATO									
TO **FROM**		**First line – Induction**	**First line – Consolidation**		**First line – Remission**			**Second line – Induction[b]**	**tMDS/AML**
	Tunnel #	**2**	**1**	**2-8**	**1**	**2-24**	**25+**	**1**	**-**
First line – Induction	**1**	54.5%	45.5%					0.0%	
	2		100.0%					0.0%	
First line – Consolidation	**1-7**			100.0%					0.0%
	8				100.0% [99.9%]			0.0% [0.1%]	0.0%
First line – Remission	**1-24**					100.0% [100.0%[a]]		0.0% [0.0%[a]]	
	25+						100.0%[a]	0.0%[a]	
AATO+AIDA									
TO **FROM**		**First line – Induction**	**First line – Consolidation**		**First line – Remission**			**Second line – Induction[b]**	**tMDS/AML**
	Tunnel #	**2**	**1**	**2-8**	**1**	**2-24**	**25+**	**1**	**-**
First line – Induction	**1**	54.5%	45.5%					0.0%	
	2		100.0%					0.0%	
First line – Consolidation	**1-7**			100.0%					0.0%
	8				98.4% [99.9%]			1.6% [0.1%]	0.0%
First line – Remission	**1-24**					100.0%[a]		0.0%[a]	
	25+						100.0% [100.0%[a]]	0.0% [0.0%[a]]	
AIDA+AATO									
TO **FROM**		**First line – Induction**	**First line – Consolidation**		**First line – Remission**			**Second line – Induction[b]**	**tMDS/AML**
	Tunnel #	**2**	**1**	**2-3**	**1**	**2-24**	**25+**	**1**	**-**
First line – Induction	**1**	56.2% [55.1%]	40.9% [42.6%]					3.0% [2.4%]	
	2		97.0%					3.0%	
First line – Consolidation	**1-2**			99.5%					0.5%
	3				94.4% [97.6%]			5.1% [1.9%]	0.5%
First line – Remission	**1-24**					99.6%		0.4%	
	25+						99.8%	0.2%	

Note the probabilities between square brackets are used in the ERG base-case

[a]This transition probability rounded to 100.0% or 0.0% (i.e. it would be less than 100.0% or more than 0.0% if more decimals would be displayed)

[b]This refers to the "second line treatment induction phase + 1 cycle consolidation" health state

Second line health states

From the "second line treatment induction phase + 1 cycle consolidation" health state patients can transit to second line consolidation (either the consolidation tunnel state within this health state if patients are still in the induction phase, or to the "second line treatment consolidation" health state), to "alloHSCT", "autoHSCT" or "tMDS/AML". The transition from second line treatment induction (first cycle) to second line treatment consolidation is based on the average induction duration[12, 39] plus the duration of one consolidation cycle (in total this resulted in an assumed median duration of 60 and 63 days for second line AATO and AIDA respectively). During the second line treatment induction phase for AIDA it is possible to transit to "alloHSCT" (not possible during second line treatment induction phase for AATO) if patients experience an adverse event requiring a treatment switch (i.e. cardiac event, identical probability as for first line AIDA treatment). From the last tunnel of the "second line treatment induction phase + 1 cycle consolidation" health state (i.e. during the second or first consolidation cycle for second line AATO and AIDA respectively) patients can transit to both HSCT health states or continue to the "second line treatment consolidation". These transitions were conditional on the haematological complete response rates of 80%[18] and 70% (expert opinion) for AATO and AIDA respectively. Due to lack of data, the company assumed that the molecular remission rate was identical to haematological complete response. Subsequently, it was assumed, that all patients without remission, that did not transit to "tMDS/AML", would transit to "alloHSCT". For the patients with remission, 10.3% and 34.5% would transit to the "alloHSCT" and "autoHSCT" health states respectively. These probabilities of transiting to the HSCT health states were based on data on file, without providing detailed information on this source.[32] The remaining patients transit to the "second line treatment consolidation" health state.

From the "second line treatment consolidation" health state patients can transit to "alloHSCT", "autoHSCT", "tMDS/AML" or "second line molecular remission". In the initial "second line treatment consolidation" tunnel states, patients can only transit to the next tunnel state or to "tMDS/AML". After the last consolidation cycle (maximum total of eight and three cycles of consolidation treatment for AATO and AIDA respectively), again 10.3% and 34.5% (based on data on file[32]) transit to the "alloHSCT" and "autoHSCT" health states respectively. The remainder, that did not transit to "tMDS/AML", transits to "second line molecular remission", implicitly assuming a 100% molecular remission rate.

The cycle probability of transiting to "tMDS/AML" from the second line AIDA treatment consolidation tunnel states was slightly higher than for the first line AIDA treatment consolidation tunnel states (0.5% versus 0.7% respectively) due to a difference in the assumed maximum duration of the AIDA consolidation phase (84 versus 56 days respectively).

From "second line molecular remission" health state, patients can transit to "alloHSCT" or "autoHSCT". It is assumed that all patients that relapse (monthly probability of 1.1%[31]) would transit to "alloHSCT". For patients without relapse, the previously mentioned probabilities of transiting to the HSCT health states (i.e. 10.3% and 34.5%, based on data on file,[32] transit to the "alloHSCT" and "autoHSCT" respectively) are adjusted using the median time to relapse (of 24.5 and 14.0 months based on first line data from the APL0406 trial[14]) after second line treatment with AATO and AIDA respectively.

The transition probabilities from the second line health states (retrieved from the model) are presented in Table 5.6. This excludes general population mortality that is subsequently applied to the transition probabilities presented in this table (no additional mortality is assumed).

ERG comment: The ERG's concerns relate to (a) evidence and transparency of descriptions for transitions from second line health states and; errors concerning the transition probability (b) from second line induction to consolidation for AIDA, (c) from second line AIDA consolidation to "tMDS/AML", as well as (d) from "second line molecular remission" to the HSCT states. The transition probabilities that are adjusted in the ERG base-case are presented between square brackets in Table 5.6.

(a) The evidence presented to inform transitions from second line health states was weak and it was frequently not transparently reported how the transition probabilities were obtained. For instance, evidence was obtained from expert opinion, based on 10 patients that received AATO[18] and first line data from the APL0406 trial was used. Moreover, it was unclear how transition probabilities were obtained from Tallman et al[31] (secondary data source that describes a calibration process to obtain transition probabilities that could not be reproduced by the ERG) and Russell et al[32] (data on file). This hampered the assessment of these transition probabilities.

(b) The ERG found a reference error in cell 'Calc - TransMat ATRA+ATO'!AT141 (cell refers to J10 while it should be J96, in the economic model initially submitted by the company). This affects the transition from second line induction to consolidation for AIDA.

(c) For calculating the probability of transiting to "tMDS/AML" during second line AIDA consolidation treatment, a maximum consolidation duration of 56 days (i.e. two cycles) is assumed while the probability is applied during 84 days (three cycles). This is corrected in the ERG base-case.

(d) The transition from "second line molecular remission" to the HSCT states is adjusted using the median time to relapse following second line remission. The rationale of this adjustment is unclear to the ERG. However, the uncorrected transition probabilities to the HSCT states seem relatively high (given these are applied each cycle). Hence, the ERG explored using the uncorrected transition probabilities to the HSCT states as well as setting these to 0.0% in scenario analyses.

Table 5.6: Transition probabilities (per cycle) from the second line health states

Second line AATO (re)treatment										
TO **FROM**		**Second line – Induction + 1 Consolidation[a]**			**Second line – Consolidation**		**Second line – Remission**	**AlloHSCT**	**AutoHSCT**	**tMDS/AML**
	Tunnel #	**2**	**3[b]**	**4[c]**	**1**	**2-6**	**-**	**1**	**1**	**-**
Second line – Induction + 1 Consolidation	**1**	72.4%	27.6%					0.0%		
	2		100.0%					0.0%		
	3[b]			100.0%						
	4[c]				44.2%			28.2%	27.6%	0.0%
Second line – Consolidation	**1-5**					100.0%				0.0%
	6						55.2%	10.3%	34.5%	0.0%
Remission	**-**						96.7%	1.5%	1.7%	
Second line AIDA treatment (after first line AATO)										
TO **FROM**		**Second line – Induction + 1 Consolidation**			**Second line – Consolidation**		**Second line – Remission**	**AlloHSCT**	**AutoHSCT**	**tMDS/AML**
	Tunnel #	**2**	**3[b]**	**4[c]**	**1**	**2**	**-**	**1**	**1**	**-**
Second line – Induction + 1 Consolidation	**1**	69.4% [71.1%]	27.6% [26.5%]					3.0% [2.4%]		
	2		97.0% [97.6%]					3.0% [2.4%]		
	3[b]				38.6% [38.6%]			36.5% [36.7%]	24.2%	0.7% [0.5%]
	4[c]									
Second line – Consolidation	**1**					99.3% [99.5%]				0.7% [0.5%]
	2						54.5% [54.7%]	10.3%	34.5%	0.7% [0.5%]
Remission	**-**						95.1%	1.9%	3.0%	

Note the probabilities between square brackets are used in the ERG base-case
[a]For AATO (re)treatment this health state contains 2 consolidation cycles (tunnel states 3 and 4)
[b]This tunnel state represents the first consolidation cycle
[c]This tunnel state represents the second consolidation cycle (not used for second line AIDA treatment)

Haematopoietic stem cell transplant and other health states

This section considers the haematopoietic stem cell transplant as well as "tMDS/AML" and "death" health states in the model.

From the "alloHSCT" health state patients can transit to "molecular remission after alloHSCT", "end of life" (health state for patients that failed on HSCT) and "death". In the initial "alloHSCT" tunnel states, patients can only transit to the next tunnel state or to "death". After the last "alloHSCT" tunnel state (i.e. after six tunnel states; 168 days), patients can transit to "molecular remission after alloHSCT", "end of life" and "death". The death probability is identical (i.e. 6.7% per cycle[33]) as in the previous "alloHSCT" tunnel states. For the surviving patients it is assumed that 72.2%[36] would have a molecular remission and hence transit to "molecular remission after alloHSCT" while the remaining patients transit to the "end of life" health state.

Similar to the transitions from alloHSCT, from the "autoHSCT" health state patients can transit to "molecular remission after autoHSCT", "end of life" and "death". In the initial "autoHSCT" tunnel states, patients can only transit to the next tunnel state or to "death". After the last "autoHSCT" tunnel state (i.e. after three tunnel states; 84 days), patients can transit to "molecular remission after autoHSCT", "end of life" and "death". The death probability is identical (i.e. 2.0% per cycle[33]) to the previous "alloHSCT" tunnel states. For the surviving patients it is assumed that 98.1%[36] would have a molecular remission and hence transit to "molecular remission after alloHSCT" while the remaining patients transit to the "end of life" health state.

From the "molecular remission after alloHSCT" and "molecular remission after autoHSCT" health states patients can only transit to "death". The transition probabilities are 0.2%[35] per cycle for both health states. Similarly, from the "end of life" and "tMDS/AML" health states, patients can only transit to "death" as well. The transition probabilities were 3.1%[34, 35] and 2.7%[37] per cycle for the "end of life" and "tMDS/AML" health states respectively.

The transition probabilities from the abovementioned health states (retrieved from the model) are presented in Table 5.7. This excludes general population mortality that is subsequently applied to the transition probabilities presented in this table.

ERG comment: The ERG's concerns relate to a general lack of (a) detailed descriptions and justification for calculations, assumptions and selected sources and; (b) elaborate consideration in the CS of (implicit) assumptions regarding extrapolation.

(a) Most of the evidence sources mentioned above are not described in the CS (neither are the transition probabilities). Although this is most prominent for the transition probabilities described in the preceding section, the lack of detailed description and justification in the CS is applicable to the majority of the transition probabilities described in the treatment effectiveness section of the ERG report. This is worrying, given treatment effectiveness (including implicit assumptions made and selection of evidence sources to obtain transition probabilities) is often an influential part of the cost effectiveness model. Although the company's response to clarification question B6 was helpful, justification for calculations, assumptions and selected sources remain largely unclear to the ERG.

(b) The extrapolation of treatment effectiveness is a limitation of the model (as also indicated by the company in response to clarification question B7) that is not extensively discussed in the CS. This includes for instance the extrapolation of relative treatment effectiveness, i.e. implicitly assuming that treatment benefits are maintained for the entire time horizon. In the first line for example, the relapse transition probability is assumed to be constant from two years after remission until the end of the time

horizon (higher for AIDA than for AATO). Alternative assumptions regarding the extrapolation could be influential as illustrated by the analyses performed by the company in response to clarification question B7. In this scenario, assuming no relapse after 24 months in the first line remission health state, AATO did not remain dominant as it became more expensive than AIDA with an ICER of £7,610 per QALY gained. Acknowledging this uncertainty, the ERG added a scenario analyses assuming an equal relapse probability two years after first line remission for AATO and AIDA.

Table 5.7: Transition probabilities (per cycle) from the haematopoietic stem cell transplant or tMDS/AML health states

TO FROM		AlloHSCT	AlloHSCT – Remission	AutoHSCT	AutoHSCT – Remission	End of life	tMDS/AML	Death[a]
	Tunnel #	2-6	-	2-3	-	-	-	-
AlloHSCT	1-5	93.3%						6.7%
	6		67.4%			25.9%		6.7%
AlloHSCT – Remission	-		99.8%					0.2%
AutoHSCT	1-2			98.0%				2.0%
	3				96.1%	1.9%		2.0%
AutoHSCT – Remission	-				99.8%			0.2%
End of life	-					96.9%		3.1%
tMDS/AML	-						97.3%	2.7%

[a]Disease related mortality

5.2.7 Adverse events

The following treatment-induced adverse events were considered in the model:

- Thrombocytopenia (grade 3–4, duration >15 days)
- Neutropenia (grade 3–4, duration >15 days)
- Infection
- Leukocytosis
- Hepatic toxicity
- Neurotoxicity
- Differentiation syndrome
- Cardiac events
- QTc prolongation

Except for cardiac events, the adverse events listed above did not lead to a change of treatment, but impacted only the costs and patients' QoL. The duration of each adverse event was used to compute the QALYs lost due to the QoL impairment in patients experiencing the event.

In addition to the adverse events listed above, tMDS or AML was incorporated in the model structure (see section 5.2.2).

ERG comment: Similar to the treatment effectiveness parameters, the description of adverse events (AEs) in the CS lacked transparency. The AE probabilities were not mentioned in the CS. See Table 5.8 for an overview that was retrieved from the model submitted by the company (see section 5.2.6 for the probability of therapy-induced MDS or AML). Although the company's response to clarification question B6 was helpful, justifications for the selected sources are largely unclear to the ERG. This was particularly the case for the sources selected to inform the duration of AEs, including multiple sources that are published >25 years ago. Moreover, the selection of these specific AEs is unclear to the ERG; this includes that is was unclear why reversible arrhythmia was not considered in the model (as discussed in section 5.2.6).

Table 5.8: Adverse events used in the economic model (for both first line and second line AATO/AIDA treatment)

	AATO		AIDA	
	Induction	**Consolidation**	**Induction**	**Consolidation**
Thrombocytopenia (grade 3-4, >15 days)				
Probability	0.380[14]	0.187[14]	0.620[14]	0.810[14]
Duration (days)	20[40]	25[40]	20[40]	25[40]
Neutropenia (grade 3-4, >15 days)				
Probability	0.350[14]	0.127[14]	0.640[14]	0.813[14]
Duration (days)	19[41]	19[41]	19[41]	19[41]
Infection				
Probability	0.230[14]	0.042[14]	0.550[14]	0.152[14]
Duration (days)	17[42]	17[42]	17[42]	17[42]
Leukocytosis				
Probability	0.473[12]	0.000[12]	0.241[12]	0.000[12]
Duration (days)	14[43]	14[43]	14[43]	14[43]
Hepatic toxicity (grade 3-4)				
Probability	0.400[14]	0.016[14]	0.030[14]	0.002[14]
Duration (days)	10[44]	10[44]	10[44]	10[44]
Neurotoxicity (all grades)				

Probability	0.007[14]	0.050[14]	0.000[14]	0.000[14]
Duration (days)	365[45]	365[45]	365[45]	365[45]
Differentiation syndrome				
Probability	0.194[12]	0.000[12]	0.160[12]	0.000[12]
Duration (days)	4[46]	4[46]	4[46]	4[46]
Cardiac events (grade 3-4)				
Probability	0.000[14]	0.000[14]	0.037[14]	0.000[14]
Duration (days)	1[47]	1[47]	1[47]	1[47]
QTc prolongation				
Probability	0.085[14]	0.018[14]	0.007[14]	0.000[14]
Duration (days)	0.5[48]	0.5[48]	0.5[48]	0.5[48]

5.2.8 Health-related quality of life

Health state utility values

Both the APL0406 and the AML17 trials used the EORTC QLQ-C30 instrument, and not the EQ-5D, to measure HRQoL outcomes. Therefore, the company performed a SLR to identify relevant quality of life studies for the current decision problem which yielded two CEA studies focussing on APL patients.[31, 38] The study by Tallman et al. 2015[31] used utility values from chronic lymphocytic leukaemia (CLL) patients which they adjusted for age and country (adjustment method not described in the CS). Lachaine et al. 2015[38] used utility values from AML patients. The CS did not report the primary sources informing the utility values in these CEA studies. No study reporting utility values based on the EQ-5D for APL patients was identified in the literature (see section 5.1 for more details regarding the SLR).

Utility values used in the model were obtained from the study by Woods et al. 2012,[49] which reported utility values from CLL patients. The company states that this study was selected because it "presented utility values for similar health states to those in our model, reflecting the treatment pathway".[1] Woods et al. 2012[49] was presumably identified through the study by Tallman et al. 2015.[31] Additionally, the company used Beusterien et al. 2010[50] (also considering CLL patients) because it was referred to in Woods et al., 2012.[49] Szende et al. 2014[51] provided general population utility values for the current assessment. The CS did not describe how the utility values were obtained in these studies (Woods et al., 2012[49], Beusterien et al. 2010[50] and Szende et al. 2014[51]) and why these sources were deemed to be the most appropriate for the current decision problem.

The company adjusted the utility values obtained from the literature, with the intention to make them more relevant for the modelled population. Two adjustments were made: 1) an adjustment for age and; 2) an adjustment for the utility representing perfect health:

1. The age adjustment consisted of multiplying with the ratio of the utility in the general population having the same age as the modelled population (i.e. 45 years old) to the utility value in the general population with the same age as the population in which the utility value was obtained (e.g. 60 years old).[1] The UK general utility values for patients aged 45 and 60 are 0.849 and 0.804 respectively, which resulted in a factor of 1.056 (=0.849/0.804) for the age adjustment.
2. The adjustment for the utility representing perfect health consisted of multiplying the (age-adjusted) utility values by the utility value in the UK general population with the same age as the population in which the utility was elicited. This adjustment for a 60 year old patient population would then be 0.804 (i.e. UK general population utility value of a 60 year old person).

When applying both adjustments to a utility value 0.910 obtained from the literature for a 60 year old patient population, these adjustments would result in a utility value of 0.773 (= 0.910×1.056

(=0.849/0.804) × 0.804) for a 45 year old patient in the modelled population. These two adjustments combined effectively equal the multiplication by 0.849 (i.e. the general population utility value of the 45 year old modelled population). It should be noted that the company did not apply these adjustments consistently on all health state utility values obtained from the literature (see Table 5.9 for details regarding the application of the adjustments). No evidence was provided to support the need to adjust the original utility values and no justification supported the inconsistencies in adjustments between health states.

The company applied another adjustment for patients receiving second line treatment (second line induction and consolidation phase). Here, the second line treatment utility was assumed to be 91% of the first line treatment utility value. The 91% was calculated by dividing 0.71 by 0.78, which represented utility values of Stable CLL during second and first line treatment respectively (utility values presumably obtained from Woods et al., 2012[49]). No evidence was provided in the CS to justify this adjustment.

Finally, an adjustment was made to obtain the utility values for the "Allogeneic HSCT" and "Autologous HSCT" health states: the utility value of the "CML after HSCT without GvHD" (i.e. 0.979) was multiplied by the utility in the "second line molecular remission" state (i.e. 0.702), which resulted in 0.687 (Table 5.9). The primary source for the "CML after HSCT without GvHD" utility value and the rationale for adjusting it were not provided.

Patients in the long-term remission health states were assumed to have a utility value equal to the general population at the age of 45. Table 5.9 presents the utility values from the original source, the different adjustments and the utility values used in the cost effectiveness model.

Table 5.9: Overview of the health state utility values used in the model

Health state	Mean utility value in cost effectiveness model (95% CI)	Age adjustment[a]	Adjustment for general population utility value[a]	Original utility value (95% CI)	Adjustment for second line treatment	Disutility for hospitalisation
First line induction treatment	0.74 (0.71, 0.77)	1.06	None	0.70[b] (0.67, 0.73)	NA	-0.01[g]
First line consolidation treatment	0.74 (0.71, 0.77)	1.06	None	0.70[b] (0.67, 0.73)	NA	No
First molecular remission	0.77 (0.75, 0.79)	1.06	0.81	0.91[c] (0.88, 0.93)	NA	No
First long-term remission (>2 years)	0.85 (NR)	NA	NA	0.85[a] (NR)	NA	No
Second line induction + 1 cycle consolidation	0.67 (0.64, 0.70)	1.06	None	0.70[b] (0.67, 0.73)	0.91[c]	-0.01[g]
Second line treatment consolidation	0.70 (NR)	None	None	0.77[c,h] (NR)	0.91[c]	-0.01[g]
Second molecular remission	0.85 (NR)	NA	NA	0.85[a] (NR)	None	No
Allogeneic HSCT*	0.69 (NR)	None	None	0.98[d] (NR)	None	-0.01[g]
Autologous HSCT*	0.69 (NR)	None	None	0.98[d] (NR)	None	-0.01[g]
Allogeneic HSCT molecular remission	0.85 (NR)	None	None	0.85[a] (NR)	None	No
Autologous HSCT molecular remission	0.85 (NR)	None	None	0.85[a] (NR)	None	No
End of life state (Palliative care)	0.40[f] (NR)	None	None	0.40[e] (NR)	None	No
tMDS/AML	0.40[g] (NR)	None	None	0.40[f] (NR)	None	No

Source: Adapted from CS, Table 3.4

* The utility weight in 'CML after HSCT without GvHD' (i.e. 0.979) was adjusted by the utility in the "second line molecular remission" state (i.e. 0.702): 0.979×0.702= 0.687.

[a] Age-adjustments, adjustments for general population utility values, and general population utility values were based on Szende et al., 2014[51]

[b] Obtained from Woods et al., 2012[49]; [c] Obtained from Beusterien et al., 2010[50]; [d] Obtained from Breitscheidel L., 2008[52]; [e] Obtained from Morton et al., 2009[53]; [f] Obtained from Cooperberg et al., 2013[54]; [g] Assumption

[h] In the ERG base-case, this utility value is assumed to be equal to the "Second line induction + 1 cycle consolidation" health state utility value (i.e. 0.70)

Abbreviations: CI, confidence interval; HSCT, hematopoietic stem cell transplantation; NA, not applicable; NR, not reported; tMDS/AML, treatment-related myelodysplastic syndrome or acute myeloid leukaemia.

Adverse events

Disutilities for adverse events were included in the treatment induction and consolidation health states of the cost effectiveness model. Adverse events were assumed not to occur in the remission health states. The duration (see Section 5.2.7 for more details) and disutilities associated with adverse events were obtained from the literature. Disutilities associated with adverse events were calculated for each cycle and were obtained by multiplying the proportion of patients experiencing the adverse event by the duration of the adverse event and the disutilities associated with the adverse events. The proportion of patients experiencing each adverse event was not reported in the CS (Section 5.2.7). The CS did not report how the sources informing the disutilities were identified and did not justify why these sources were the most appropriate.

Besides disutilities specifically applied to the treatment induction and consolidation health states, patients experienced a disutility for hospitalisation (i.e. -0.01) during first and second line treatment (both induction and consolidation) as well as during HSCT treatment. Additionally, patients in the "Allogeneic HSCT" and the "Allogeneic HSCT molecular remission" health states were at risk of experiencing a disutility for graft versus host disease (GvHD). The proportion of patients experiencing GvHD, the duration of GvHD, and the duration of hospitalisation in the above-mentioned health states were not reported in the CS. The CS also emphasised that patients could experience acute and chronic GvHD but did not describe how these were differentiated in the cost effectiveness model. Table 5.10 provides an overview of the duration and disutility values associated with adverse events.

Table 5.10: Overview of adverse events duration and associated disutilities

State	Mean utility value	Reference	Justification (comment)	Mean duration	Reference	Justification (comment)
Hospitalisation	-0.01	Assumption				
Thrombocytopenia (grade 3-4, >15 days)	-0.18	Attard et al., 2014[55]		Induc.: 20 days Cons.: 25 days	Wolff et al., 1989[40]	
Neutropenia (grade 3-4, >15 days)	-0.18	Attard et al., 2014[55]		19 days	Fenaux et al., 1993[41]	Assumed to be the same as for ATRA+DNR+ARA-C
Infection	-0.15	Stevenson et al., 2014[56]	Based on table A1 in Platzbecker 2016[3], most infections are pneumonia. Disutility of pneumonia was considered.	17 days	Pneumonia – What happens[42]	Based on table A1 in Platzbecker et al. 2016[3], most infections are pneumonia, thus duration of pneumonia (2-3 weeks) was considered.
Leukocytosis	-0.08	Assumption		14 days	Shoenfeld et al., 1981[43]	
Hepatic toxicity	-0.2	Choi et al., 2013[57]		10 days	Zhu et al., 2013[44]	“Less than two weeks”
Neurotoxicity	-0.21	Prica et al., 2014[58]		365 days	Assumption based on Ratnaike, 2003[45]	"Acute poisoning from arsenic can lead to peripheral neuropathy which can last for max 2 years"
Differentiation syndrome	-0.12	Assumption		4 days	Breccia et al., 2008[46]	Assumed to be same as AIDA
Cardiac events	-0.16	Nshimyumukiza et al., 2013[59]	· Myocardial infarction (MI)	1 day	Mathews et al., 2002[47]	Assumed to be same as ATRA+ATO
QTc prolongation	-0.001	Assumption		0.5	Siu et al., 2006[48]	Assumed to be same as ATRA+ATO
Acute GvHD	-0.08	Breitscheidel L., 2008[52]	· Mean utility weight after HSCT without GvHD, re-scaled: 0.836	NR		

State	Mean utility value	Reference	Justification (comment)	Mean duration	Reference	Justification (comment)
			· Mean utility weight after HSCT with GvHD, re-scaled: 0.769	NR		
			· Disutility of GvHD: 1-(0.769/0.836)= 0.080	NR		
			· Applied for the duration of the monitoring phase for the proportion of patients experiencing acute GvHD	NR		
Chronic GvHD	-0.08	Breitscheidel L., 2008[52]	Assumed to be the same as acute GvHD	NR		
			Applied for a lifetime for the proportion of patients experiencing chronic GvHD	NR		

Source: Adapted from CS, Table 3.4

Abbreviations: AIDA, all-trans retinoic acid and idarubicin; Ara-C, cytarabine; ATO, arsenic trioxide; ATRA, all-trans retinoic acid; Cons., consolidation; DNR, daunorubicin; GvHD, graft versus host disease; Induc., induction; NR, not reported.

ERG comment: The ERG agrees with the company that utility values for APL patients elicited through the EQ-5D are likely not to be available in the literature. However, the ERG is concerned by the validity of the utility values for the following reasons: (a) the selection process of the utility values and the assumptions underlying disutilities associated with adverse events were unclear, (b) the non-adherence with the NICE reference case, and (c) the lack of justification supporting the adjustments made by the company.

(a) In the CS, the company refers to Woods et al. 2012, Beusterien et al. 2010 and Szende et al. 2014 as main sources informing health state utility values. Breitscheidel 2008[52], Morton et al. 2009[53], and Cooperberg et al. 2013[54] were also used to inform the utility values of the "End of life state (Palliative care)", "tMDS/AML", and "Allogeneic HSCT" and "Autologous HSCT" health states, respectively (Table 5.9). Moreover, additional sources and/or assumptions were used to inform the disutility values associated with adverse events. However, the CS did not describe how all these studies were identified and justifications or evidence supporting the assumptions made by the company were largely lacking. Therefore, the ERG requested that the company clarify the choice of different sources and assumptions made in the CS. The company responded that a targeted search was performed to identify health state utility values and disutility values associated with adverse events. The selection of utility and disutility values and the assumptions supporting the disutility values associated with adverse events had then been discussed with experts. No detail was provided on the targeted literature search (e.g. key words used to identify studies and databases which were included in this search) and the expert opinion elicitation method. Hence, the ERG was not able to assess the quality of the selection process and thus it is unclear whether the most appropriate sources have been used to inform the utility values.

(b) The selected utility values do not adhere to the NICE reference case because utility values have not been directly elicited in patients affected by APL through the EQ-5D. The CS refers to Woods et al. 2012[49] as the primary source for the utility values of several health states (Table 5.9). Woods et al. 2012 is a cost effectiveness analysis of bendamustine versus chlorambucil for the first line treatment of CLL in England and Wales.[49] Utility values in this study were obtained by mapping European Organisation for Research and Treatment of Cancer C30 quality of life data elicited in CLL patients (collected by Knauf et al.) to EQ-5D utility values. Woods et al. 2012 [49] does not provide a precise reference to the utility elicitation study of Knauf et al.. Besides Knauf et al., Woods et al. 2012[49] also used utility increments and decrements from Beusterien et al. 2010,[50] which is a vignette study in 89 members of the general population. In this study, the standard gamble technique was used to elicit utility values associated with different health states observed during CLL treatment. It was unclear to the ERG why CLL patients were the most appropriate proxy to obtain utility values for the current decision problem. Finally, Breitscheidel, 2008[52], Morton et al. 2009[53], and Cooperberg et al. 2013[54] are cost effectiveness analyses in chronic myeloid leukaemia, kidney transplant recipients, and prostate cancer patients, respectively.

(c) Although the ERG agrees that the utility values identified in the literature are not completely reflecting the population considered in the current decision problem, it is highly questionable whether the adjustments applied by the company improves the validity of the utility values. Firstly, the necessity of an age-adjustment is questionable since the impact of the disease on quality of life would outweigh the impact of age-related utility decrements. Secondly, the ERG thinks that the adjustment for the utility representing perfect health should not be applied because no evidence (nor justification) was provided to support the methodology used by the company. The company states that this adjustment ensures that the health state utility values would not be higher than the general population utility values. However, this is only true for the start of the model, i.e. not over time, which is a severe limitation given the life expectancy of the modelled patients. Thirdly, these adjustments were not applied consistently on all

health states (and the rational for doing so is missing); hence the ERG decided not to use the company's adjustment in its base-case analysis (health state utilities in column 'Original utility value' from Table 5.9 are used in the ERG base-case). In order to prevent that health state utility values exceed the general population utility values (over time), the ERG decided to cap the health state utility values in the model using the general population utility values (see scenario requested in clarification question B11).[60] Additionally, the ERG decided to use the same utility value for the first and second line treatment induction and consolidation health states (i.e. 0.70). This was adopted to ensure consistency in utility values between these health states because the company did not provide evidence to support the need for differential utility values.

5.2.9 Resources and costs

The cost categories included in the model were treatment acquisition costs, medical costs (treatment administration, supportive care, monitoring and follow-up, HSCT, palliative care), and costs of managing adverse events.

Treatment acquisition costs

The company states that publications of relevant trials were used to extract data on dosage and number of doses of intervention and comparator, and that these were validated by clinical experts. Drug costs were based on the British National Formulary (BNF) [61]. If available, container sizes minimising the costs and wastage were used to reflect real-life practice. Input values and their sources for each drug and treatment phase are presented in Tables 5.11 and 5.12.

Table 5.11: Unit treatment acquisition costs associated with the technologies studied in the economic model

		INDUCTION PHASE			CONSOLIDATION PHASE		
Model parameter	**Strategy**	**Drug**	**Value**	**Reference**	**Drug**	**Value**	**Reference**
Number of doses	AATO First line	ATRA	32	Lo-Coco et al., 2013	ATRA	15	Lo-Coco et al., 2013[12]
		ATO	32		ATO	20	
	AATO Second line	ATRA	25	Douer et al. 2005	ATRA	15	Lo-Coco et al., 2013[12]
		ATO	25		ATO	25	SPC
	AIDA	ATRA	35	Lo-Coco et al., 2013	ATRA	15	Lo-Coco et al., 2013[12]
					IDA(cycle 1)	4	
					Mitoxantrone (cycle 2)	5	
		IDA	4		IDA(cycle 3)	1	
Indicated dose per day	AATO	ATRA	45 mg/m^2	Lo-Coco et al., 2013	ATRA	45 mg/m^2	Lo-Coco et al., 2013[12]
		ATO	0.15 mg/kg		ATO	0.15 mg/kg	
	AIDA	ATRA	45 mg/m^2		ATRA	45 mg/m^2	
					IDA(cycle 1)	5 mg/m^2	

		INDUCTION PHASE			CONSOLIDATION PHASE		
Model parameter	**Strategy**	**Drug**	**Value**	**Reference**	**Drug**	**Value**	**Reference**
		IDA	12 mg/m^2		Mitoxantrone (cycle 2)	10 mg/m^2	
					IDA(cycle 3)	12 mg/m^2	
Container size	AATO	ATRA	10 mg	BNF	-	-	-
		ATO	10 mg				
	AIDA	ATRA	10 mg		Mitoxantrone	20 mg	BNF[61]
		IDA	10 mg				
Cost per container	AATO	ATRA	£1.61	BNF	-	-	-
		ATO	£292.00				
	AIDA	ATRA	£1.61		Mitoxantrone	£100.00	BNF[61]
		IDA	£174.72				

Table 5.12: Costs of technology per treatment phase

Phase	**AATO**			**AIDA**		
	ATRA	**ATO**	**Total AATO**	**ATRA**	**Chemo (IDA+MTZ)**	**Total AIDA**
First line: induction	£463.68	£16,078.58	£16,542.26	£507.15	£2,096.64	£2,603.79
First line: Consolidation	£1,521.45	£40,196.44	£41,717.89	£652.05	£1,723.04	£2,375.09
First line: Total	£1,985.13	£56,275.02	£58,260.15	£1,159.20	£3,819.68	£4,978.88
Second line: Induction	£362.25	£12,561.39	£12,923.64	£507.15	£2,096.64	£2,603.79
Second line: Consolidation	£1,521.45	£12,561.39	£14,082.84	£652.05	£1,723.04	£2,375.09
Second line: Total	£1,883.70	£25,122.77	£27,006.47	£1,159.20	£3,819.68	£4,978.88

Medical costs

Medical costs (treatment administration, supportive care, monitoring and follow-up, HSCT, palliative care), and resource use are presented in Table 5.13 below. Medical costs were mainly based on NHS reference prices, BNF and PSSRU.

Table 5.13: list of resource use per health state

Health state	**Items**	**AATO**	**Value**	**AIDA**	**References**
Induction phase	Number of bed days per patient	First line: 32 Second line: 25	£396.47	35	AATO: First line: Lo-Coco et al., 2013[12] Second line: Douer et al., 2005[39]

Health state	Items	AATO	Value	AIDA	References
					AIDA: Lo-Coco et al., 2013[12]
	Number of supportive care transfusions	15	0	22	Burnett et al., 2015[13]
	Number of annual PCR tests	5	£280	4	Expert opinion
Consolidation phase	Number of bed days per patient	0	£396.47	4	AATO: Expert opinion AIDA: assumption based on treatment schedule
	Number of ambulatory days per patient	First line: 10 Second line: 12.5	£162.00	0	AATO: Expert opinion AIDA: Inpatient treatment assumed
	Number of days of antibiotics	1	£1.65	2	Burnett et al., 2015[13]
	Number of annual PCR tests	5	£280	4	Expert opinion
Molecular remission (first, second, allo- and auto-HSCT)	Duration of follow-up	3	£210	3	First remission: Platzbecker et al., 2015[14] Others: Expert opinion
	Number of annual appointments	4	£52.50	4	First remission: Platzbecker et al., 2015[14] Others: Expert opinion
	Number of annual PCR tests	4 (0 at first remission)	£280	4	First remission: AATO: Expert opinion AIDA: Platzbecker et al., 2015[14] Others: Expert opinion
Allogeneic HSCT	Hospitalisation duration (weeks)	4	£27,907.53	4	Expert opinion
Autologous HSCT	Hospitalisation duration (weeks)	3	£7,122.97	3	Expert opinion

Costs of managing adverse events

Cost per occurrence for each type of adverse event was searched in the National Schedule of Reference Costs, 2014-2015.[62] If unavailable, recent publications reporting English or UK costs were used. In case only foreign costs could be used, these were converted to sterling using the annual exchange rates of the year the cost related to and uplifted to 2015 using inflation rates of the Office for National Statistics (Table 5.14).

Table 5.14: List of adverse reactions and summary of costs in the economic model

Adverse reactions	Value	Reference
Thrombocytopenia (grade 3-4, >15 days)	£1,746.00	NHS Reference Costs 2014–15[62]
Neutropenia (grade 3-4, >15 days)	£2,845.43	Morgan et al., 2007[63]
Infection	£253.97	Soini et al., 2016[64]
Leukocytosis	£349.44	Expert opinion
Hepatic toxicity	£5.56	Akhtar and Chung, 2014[65]
Neurotoxicity	£675.88	Calhoun et al., 2001[66]
Differentiation syndrome	£1,225.23	Milligan et al., 2006[67]; BNF[61]; National Schedule of Reference Costs[62]
Cardiac events	£1,104.02	National Schedule of Reference Costs[62]
QTc prolongation	£34.50	Expert opinion; NICE clinical guideline, No. 108[68]; NICE clinical guidelines, No. 174[69]
tMDS/AML	£6,207.00	National Schedule of Reference Costs[62]
Acute GvHD	£34,493.05	Saito et al., 2008[70]
Chronic GvHD	£8,785.25	Jones et al., 2016[71]

ERG comment: The ERG comments are in relation to (a) justification of sources for cost and resource items and (b) absence of costs related to haematological response monitoring.

(a) The ERG asked the company to provide more specific justification for each resource use and cost item. The company responded that they aimed to use NHS reference costs and the PSSRU wherever possible, supplementing this with data from studies identified through a targeted search where necessary. However, the company could not provide further justification and details about the included targeted sources, and the ERG was therefore unable to assess whether these sources were best available evidence to inform resource use and costs estimates.

(b) Costs related to monitoring of haematological response were not included in the model. The company stated that this was not mentioned by experts, most likely because the benefits of treating a molecular relapse before progression into a haematological relapse are widely recognised, so the monitoring for relapse focuses more on molecular (PCR) testing. Furthermore, the company considered the cost of haematological monitoring negligible compared to the costs of PCR testing. The ERG agrees that in first line treatment monitoring costs would be equivalent in both strategies. However, given the fact that AIDA patients relapse more frequently in second line, less monitoring would be needed in this strategy, and therefore monitoring costs would be higher in the AATO strategy.

5.2.10 Cost effectiveness results

In the deterministic base-case analysis, total QALYs and LYs gained were larger in the AATO strategy compared to the AIDA strategy. This was mostly explained by the 10.72% increase of patients in first molecular remission and the absence of tMDS/AML in the AATO arm. Furthermore, the number of APL-related deaths was reduced by around 31.85% with AATO compared to AIDA. Total costs were lower for AATO, thus the combination of AATO was dominant and no base-case ICER was calculated. Most important cost driver for AATO was treatment acquisition costs, but costs in all other categories were higher for AIDA. Transplantation costs were the highest costs for AIDA. Base-case health outcomes, discounted costs and the incremental cost-effectiveness ratio are shown in Tables 5.15-5.17.

Table 5.15: Discounted health outcomes in the model

	AATO	AIDA	AATO vs. AIDA
Number of QALYs	16.34	13.72	2.62
Number of LYs	19.56	16.56	3.00
First remission*	99.83%	89.11%	10.72%
First long remission (> 2 years)*	92.84%	76.11%	16.73%
MDS*	0.00%	1.39%	-1.39%
Death* (not discounted)	57.04%	74.13%	-17.09%
APL related death* (not discounted)	7.54%	39.38%	-31.85%
Background death* (not discounted)	49.51%	34.75%	14.76%

AATO= arsenic trioxide + all-trans retinoic acid; LY=life years; MDS=myelodysplastic syndrome; QALY=quality-adjusted life years
*proportion of patients ever, transiting to this health state

Table 5.16: Discounted disaggregated and total costs – base-case scenario

Cost category	AATO	ATRA+AIDA	AATO vs. AIDA
Treatments	£60,336	£21,604	£38,731
Administration	£25,402	£31,660	-£6,259
Supportive care and antibiotics	£3,575	£6,487	-£2,912
Follow-up and monitoring	£2,991	£10,389	-£7,398
Adverse Events	£4,142	£12,378	-£8,236
MDS	£0	£226	-£226
HSCT	£7,645	£48,326	-£40,681
Palliative care	£906	£5,196	-£4,290
Total	£104,996	£136,267	-£31,270

AATO= arsenic trioxide + all-trans retinoic acid; HSCT=haematopoietic stem cell transplantation; MDS=myelodysplastic syndrome

Table 5.17: Base-case incremental results (discounted)

Treatment	Total costs (£)	Total LYG	Total QALYs	Incremental costs (£)	Incremental LYG	Incremental QALYs	ICER versus baseline (£/QALY)	ICER incremental (£/QALY)
AATO	104,996	19.56	16.34	-£31,270	3.00	2.62	Dominant	Dominant
AIDA	136,267	16.56	13.72				-	-

ERG comment: The undiscounted LYs and QALYs provided by the company were 33.22 and 27.91 for AATO and 26.84 and 22.38 for AIDA, respectively. The ERG perceives the life expectancy in the model to be relatively long. This might be related to a lack of disease-specific mortality in the first and second line health states, as well as assumptions concerning (extrapolation of) treatment benefits. More details can be found in section 5.2.12 about model validation and face validity.

5.2.11 Sensitivity analyses

The company performed and presented a probabilistic sensitivity analysis (PSA) and deterministic sensitivity analyses (DSA) in order to quantify the uncertainty surrounding the base case results.

The PSA with 1,500 Monte Carlo simulations showed similar incremental costs and QALYs compared with the deterministic results, the AATO strategy still being dominant (Table 5.18). The cost effectiveness acceptability curve showed that the probability of AATO being cost effective at a willingness-to-pay (WTP) of £0 per QALY was 81%. This probability increased to 94% at a WTP of £30,000 per QALY (Figure 5.2).

The company conducted a one-way DSA to study the impact of varying individual parameter values on incremental costs, incremental QALYs and ICER of AATO compared with AIDA. Parameters that most affected incremental costs were the probability of relapse at 48 months after first remission in the AIDA arm, discount rate for costs, time horizon, complete haematological remission rate following AIDA in first line, and the probability of relapse at 24 months after first remission in the AIDA strategy. Incremental effectiveness was mostly affected by changes in the discount rate for health outcomes, time horizon, probability of relapse observed at 48 months for AIDA, first line haematological remission rate associated with AIDA treatment and the utility value in the first molecular remission (2> years) health state (Table 5.19). The ICER could only be computed in four sensitivity analyses, in all other analyses AATO dominated AIDA. The ICER based on a time-horizon of five years was 148,179 £/QALY. This can mostly be explained by the high treatment acquisition costs in the first year and the inability to capture the full HRQoL benefits within a time horizon this short.

The following scenario analyses were performed by the company (Table 5.20):

- Scenario 1: AIDA used in second line following both first line treatments
- Scenario 2: Utilities from Tallman et al.[31]
- Scenario 3: Societal perspective
- Scenario 4: AML17 protocol: a scenario using the schedule, dosage, efficacy and safety inputs based on the AML17 clinical study.
- Scenario 5: “Worst-case” scenario: a scenario accumulating unfavourable inputs for the AATO strategy.
- Scenario 6: Probability of undergoing HSCT reflecting clinical practice, with a lower proportion of patients undergoing autologous HSCT and allogeneic HSCT reserved for patients who did not achieve molecular remission after second line induction.

The ICER was dominant across all scenarios. Scenario 4 had the largest impact on both incremental costs (£66,384) and QALYs (3.39).

Table 5.18: Probabilistic sensitivity analysis results

	Incremental costs (£)	Incremental QALYs
Mean	-31,088	2.546
Median	-28,654	2.435
Min	-169,499	-8.570
Q 0.025	-110,732	-1.746
Q 0.975	32,992	6.771
Max	109,569	14.167

Figure 5.2: Cost effectiveness acceptability curve

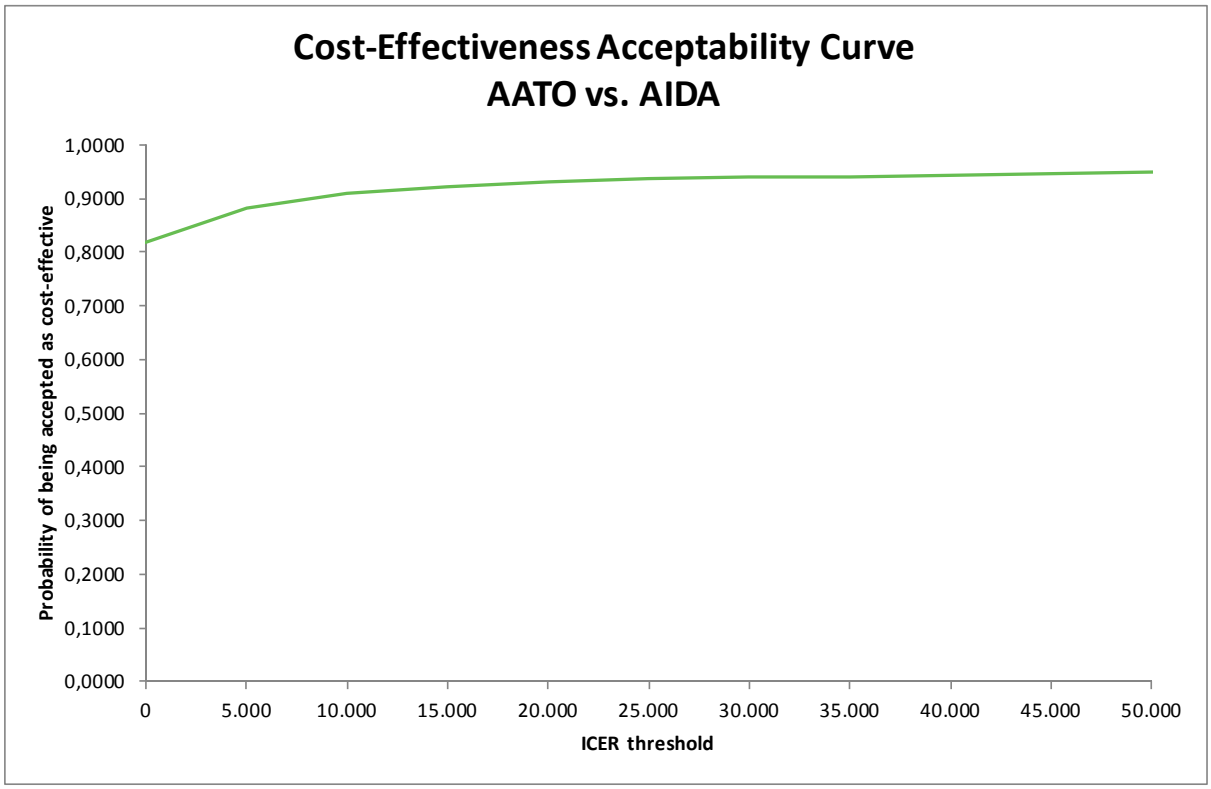


Table 5.19: Results of the DSA – incremental costs and QALYs

Parameters	**Incremental Costs (£)**		
	Low case	**High case**	**Distance***
Relapse following remission (48 months) - AIDA	25,701	-66,546	92,246
Discount rate for costs	-79,401	-19,602	59,799
Time Horizon	22,128	-25,208	59,461
CHR rate - First line - AIDA	-85,568	-27,157	58,411
Relapse following remission (24 months) - AIDAAIDA	-51,405	-510	50,896
Parameters	**Incremental QALYs**		
	Low case	**High case**	**Distance***
Discount rate for health outcomes	5.52	1.98	3.54
Time Horizon	0.15	2.07	3.02
Relapse following remission (48 months) - AIDA	1.00	3.78	2.78
CHR rate - First line - AIDA	4.84	2.45	2.39

Utilities - First line - Molecular remission (>2 years)	1.27	3.58	2.31
* Distance is ABS(Low case – Base case)+ABS(High case – Base case)			

Table 5.20: Results of scenario analyses – incremental costs, effectiveness and ratio

	AATO vs. AIDA					
	Incremental costs		**Incremental effectiveness**		**Incremental ratio**	
Scenario's	**Not discounted**	**Discounted**	**Not discounted**	**Discounted**	**Not discounted**	**Discounted**
Base-case	-	-£31,270	-	2.62	-	Dominant
Scenario 1	-£65,974	-£21,593	5.70	2.72	Dominant	Dominant
Scenario 2	-	-	6.03	2.93	-	-
Scenario 3	-	-£32,833	-	-	-	-
Scenario 4	-£125,336	-£66,384	7.10	3.39	Dominant	Dominant
Scenario 5	-£53,471	-£9,986	3.52	1.58	Dominant	Dominant
Scenario 6	-£76,110	-£28,664	5.20	2.43	Dominant	Dominant

ERG comment: The ERG concerns are related to a) the inclusion of patient characteristics and b) the approach to incorporate resource use in the PSA.

a) The ERG had minor concerns regarding the inclusion of patient characteristics (percentage male, age of patients, average height) in the PSA. Given that these parameters reflect first order uncertainty which should not be incorporated in the PSA. This is corrected in the ERG base-case.

b) The company's approach to incorporate resource use in the PSA using a normal distribution, generating new random numbers in case the resource use was negative (see response to clarification question B19b for more details), is flawed as removing these negative numbers (i.e. lower part of the distribution) will artificially increase the average of the distribution. However, given that using a Gamma distribution for resource use does not substantially influence the outcomes (see response to clarification question B19b), the ERG did not alter this in its base-case.

5.2.12 Model validation and face validity check

Internal validation

Internal validation was performed to identify programming errors, data entry issues and logical inconsistencies in the model. For this purpose, a variety of extensive tests were performed considering the following aspects of the model: efficacy and safety of compared strategies, treatment schedules, treatment costs, resource use and mortality in the modelled population. See CS Table 3.27 for a summary of the internal validation. In addition, data, calculations and formulae were verified by a person not involved in the initial project.

This part of the validation included quality control conducted following the methodology proposed by the York Health Economics Consortium (YHEC). A summary of evidence on the internal validity of the model is reported in CS Table 3.27.

External validation

An external validation was conducted, comparing the outcomes from the model to those observed in clinical trials at different time points (24 and 50 months):

- DFS in terms of proportion of patients in first remission health states (molecular remission and +2y remission).
- DFS in terms of proportion of patients in all remission health states (molecular remission, +2y remission, second line molecular remission and HSCT remission)
- OS estimated as the proportion of patients alive at a given time point in the model

None of the "absolute differences" between the trial and economic model exceeded 10 percentage points (CS Table 3.28). In general, the model overestimates outcomes (i.e. produces higher DFS and OS than observed at 24 and 50 months in the trial), with the exceptions of the AIDA first line treatment arm, in which DFS, in terms of proportion of patients in first line remission is underestimated (absolute difference ranged from -7.39% to -4.19%) and for the AATO, in which 50 months OS is underestimated (absolute difference: -0.11%). Similarly, the relative difference between AATO and AIDA is generally underestimated in the model (compared with the trial) except for DFS in terms of proportion of patients in first line remission. The company states that this might be due to the assumption in the model that patients experiencing cardiac events or failing to reach molecular remission after first line consolidation switch to second line.

ERG comment: The ERG's concerns relate to a) the lack of detailed descriptions and justification for calculations, assumptions and selected sources; b) the long life expectancy in the model; c) overestimation of proportion of patients in first line remission (illustrated in external validation); d) the lack of cross-validation and; e) the inability to perform probabilistic analyses (without errors) using the model file received in response to the clarification questions (named "ID446 arsenic trioxide TEVA CEM_v4.2_rem26 v0.1 170118 SC [noACIC].xlsm").

a) As mentioned in the preceding sections, the CS lacked transparency and appropriate justifications. This included the lack of detailed descriptions and justification for calculations, assumptions and selected sources. This also includes the lack of elaborate consideration in the CS regarding assumptions related to the extrapolation (e.g. extrapolation of treatment benefit). Currently, treatment benefits, in terms of different transition probabilities for AATO and AIDA, are maintained for the entire time horizon. Alternative assumptions regarding the extrapolation could be influential as illustrated by the analyses performed by the company in response to clarification question B7. In this scenario, assuming no relapse after 24 months in the first line remission health state, AATO did not remain dominant as it became more expensive than AIDA with an ICER of £7,610 per QALY gained. Moreover, the external validation efforts, reported in CS Table 3.28, do not consider long-term outcomes beyond 50 months. Neither did the validation section in the CS (section B3.10) include specific comments regarding the face validity of the long-term extrapolation. Hence, the long-term validity of the outcomes should be regarded as a major and potentially influential uncertainty. Acknowledging this uncertainty, the ERG added a scenario analyses assuming an equal relapse probability two year after first line remission for AATO and AIDA.

b) Related to the long-term extrapolation, the ERG perceives the life expectancy estimated in the model to be relatively long. This is likely linked to the lack of disease-related mortality in the model during the first line and second line health states (only general population mortality is considered) as well as assumptions concerning (extrapolation of) treatment benefits. The undiscounted LYs and QALYs for AATO, estimated in the model, are 33.22 and 27.91 respectively. When extending the model time horizon (to 56 years, to represent a life time horizon, which is consistent with the NICE reference case),

this would increase to 35.83 and 30.12 respectively. In the general UK population, the LY and QALYs estimated for patients aged 45 (with 48.7% being male) are 37.62 and 29.62 respectively. Hence, the outcomes estimated by the model are ~2 LYs below and ~0.5 QALYs above those for the general population.[72, 73] The latter (i.e. higher QALYs than for the general population) is likely the result of the use of utility values that exceed those for the general population over time (this is corrected in the ERG base-case). The ERG is uncertain whether these LY and QALYs, as calculated in the model, have face validity.

c) The external validation showed that the model overestimates the relative difference between AATO and AIDA (compared with the trial) for DFS in terms of proportion of patients in first line remission. This is likely related to the overestimation of the proportion of patients having cardiac events (and thus requiring a treatment switch) after AIDA treatment. This overestimation of cardiac events (illustrated in Table 2.2 of CS Appendix J) and the ERG's approach to correct this is discussed in section 5.2.6 of the ERG report.

d) As stated in section 5.2.2 of the ERG report, the company adopted a model structure that diverges from those used in other economic evaluations in this condition. Additionally, the other CEAs identified in the company's SLR resulted in positive incremental costs, while in the CS base-case AATO was cost saving. Unfortunately, the company could not perform a cross-validation to explore the exact sources for the differences in the outcomes. In response to clarification question B18 the company stated: "Since we have developed a more comprehensive approach [than previously published models], it is not possible to perform an exact cross-validation of the assumptions, inputs and outputs." The ERG believes this is reasonable.

e) The ERG was unable to perform probabilistic analyses (without errors) using the model file received in response to the clarification questions (named "ID446 arsenic trioxide TEVA CEM_v4.2_rem26 v0.1 170118 SC [noACIC].xlsm"). The ERG made its adjustments using this model file given this version of the model incorporated structural adjustments that the ERG preferred to use in its base-case. Unfortunately, given the complex implementation of the PSA in the company's model, the ERG was not able to correct the cause of this error. However, the deterministic and probabilistic results, produced by the model initially submitted by the company, are relatively similar and thus the ERG will rely on deterministic results.

5.3 Exploratory and sensitivity analyses undertaken by the ERG

Based on all considerations from Section 5.2, the ERG defined a new base-case. This base-case included multiple adjustments to the original base-case presented in the previous sections. These adjustments made by the ERG form the ERG base-case and were subdivided into three categories (derived from Kaltenthaler 2016[74]):

- Fixing errors (correcting the model where the company's submitted model was unequivocally wrong)
- Fixing violations (correcting the model where the ERG considered that the NICE reference case, scope or best practice had not been adhered to)
- Matters of judgement (amending the model where the ERG considers that reasonable alternative assumptions are preferred)

Additionally, exploratory scenario analyses were performed by the ERG to examine the potential impact of alternative assumptions on the cost effectiveness estimates (section 5.3.2). Moreover, a subgroup analysis was performed to reflect the second line population, i.e. refractory/relapsed APL (section 5.3.3).

Fixing errors

1. Number of tunnel states for the "first line molecular remission" health state (section 5.2.2).
   The ERG increased the number of tunnel states to 26 model cycles to reflect two years.
2. Overestimation of proportion of patients switching to second line induction due to experiencing a cardiac event (section 5.2.6).
   The ERG corrected the probability related to cardiac events.
3. Calculation error related to proportion of patients transiting to first line consolidation early for AIDA (section 5.2.6).
   The ERG corrected this calculation error.
4. Assumptions and calculation concerning the relapse probabilities (section 5.2.6).
   The ERG corrected errors related to the assumptions and calculation (i.e. incorrectly using unconditional probabilities as conditional probabilities as well as the lack of time correction for the 48 month relapse probability).
5. Reference error related to the transition probability from second line induction to consolidation for AIDA (section 5.2.6).
   The ERG corrected this reference error.
6. Calculation error related to transition probability from second line AIDA consolidation to "tMDS/AML" (section 5.2.6).
   The ERG corrected this calculation error.

Fixing violations

7. Time horizon not reflecting lifetime (section 5.2.5).
   The ERG extended the time horizon to 56 years to reflect a lifetime time horizon.
8. Utility adjustments (section 5.2.8).
   The ERG removed the utility adjustments made by the company and assumed a different utility for the second line consolidation phases (consistent utility as used for the other induction and consolidation phases).
9. Utility values higher than the general population utility values over time (section 5.2.8).
   The ERG capped the utility values to ensure that these would not exceed the general population utility values over time.
10. Inappropriate parameters in PSA: patient characteristics were included in the PSA (section 5.2.11).
    The ERG removed patient characteristics from the PSA.

Matters of judgment

11. Calculations and assumptions regarding the remission probability (sections 5.2.2 and 5.2.6).

    The ERG informed the remission probability based on APL0406 data and used the molecular remission rate to inform the probability of transiting to remission for patients that are evaluable with PCR (removing the additional multiplication with the haematological response rate).

Table 6.1 shows how individual adjustments impact the results plus the combined effect of all abovementioned adjustments simultaneously, resulting in the (deterministic) ERG base-case. The 'fixing error' adjustments were combined and the other ERG analyses were performed also incorporating these 'fixing error' adjustments given the ERG considered that the 'fixing error' adjustments corrected unequivocally wrong issues.

5.3.1 Deterministic ERG base-case

In the ERG base-case, incorporating all abovementioned adjustments, AATO resulted in costs savings of £23,502 and yielded 2.254 more QALYs than AIDA and hence remained dominant (see Table 6.1).

As highlighted in section 5.2.12, the ERG was unable to perform probabilistic analyses. However, the deterministic and probabilistic results, produced by the model initially submitted by the company, are relatively similar. Hence, AATO is likely to remain dominant if the ERG would be able to produce probabilistic results for its base-case.

5.3.2 Deterministic scenario analyses performed conditional on the ERG base-case

Deterministic scenario analyses were performed, conditional on the ERG base-case, to examine the potential impact of the following alternative assumptions on the cost effectiveness estimates:

12. Adding disease-related mortality, in addition to general population mortality, during the induction phases (both first and second line) using the 60 day mortality from the AML17 trial (sections 5.2.2 and 5.2.12)
13. Assuming an equal relapse probability two years after first line remission for AATO and AIDA (sections 5.2.6 and 5.2.12)
14. Replacing transitions to alloHSCT for transitions to autoHSCT (section 5.2.2)
15. Removing the transitions to the HSCT states from second line remission (section 5.2.6)
16. Assuming 'uncorrected' transitions to HSCT states from second line remission (section 5.2.6)
17. Incorporating 2% cardiac events for AATO during the induction phase, reflecting treatment switching due to potential arrhythmia (sections 5.2.2 and 5.2.6)

AATO remained dominant in these deterministic scenario analyses except for the exploratory scenario wherein the relapse probability was assumed to be equal for AATO and AIDA two years after first line remission. This scenario acknowledges uncertainty in the extrapolation of treatment benefits and hence indicates that this might be influential given this scenario resulted in an ICER of £19,734 per QALY gained (see Table 6.2 for more detailed results of the scenario analyses performed by the ERG).

Additionally, the ERG performed a worst-case scenario, implementing all scenarios analyses listed above simultaneously (except for analysis 16). This deterministic worst-case scenario resulted in an ICER of £21,622 per QALY gained.

5.3.3 Deterministic subgroup analyses performed conditional on the ERG base-case

The ERG performed a deterministic subgroup analysis, conditional on the ERG base-case, reflecting the second line population, i.e. refractory/relapsed APL (section 5.2.3). This was implemented by removing the first line health states. This analysis indicated that for this subgroup AATO would cost £18,207 more and gain 0.584 more QALYs compared with AIDA, resulting in an ICER of £31,184 per QALY gained.

5.4 Conclusions of the cost effectiveness section

The cost effectiveness searches in the company submission were all documented and reproducible. However, there were a number of inconsistencies and inaccuracies, and some redundancy. The MEDLINE and Embase search strategies used an inappropriate 'animals' limit, and it is possible that relevant evidence may have been missed as a consequence.

Although the SLR identified CEAs in the literature, the company decided to develop a de novo model. The model structure proposed by the company however diverges from those identified in the SLR. The company justified the more complex model structure by stating that the existing economic evaluations did "not adequately reflect the trajectory of APL patients". According to the company, the aim in this economic evaluation was to "offer more granularity with treatment phases, molecular remission and HSCT" to better reflect the clinical trajectory of APL patients. The ERG considers that the model

structure is appropriate to reflect this condition and treatment pathway. The economic model described in the CS is considered by the ERG to partly meet the NICE reference case. Deviations from the NICE reference case included that the population and comparators considered in the scope were not fully considered. Moreover, the HRQoL used as well as the time horizon adopted by the company deviated from the NICE reference case. The transition probabilities from the first line phase of the model were informed by the APL0406 trial. The evidence to inform transitions from second line health states was weak and it was frequently not transparently reported how the transition probabilities were obtained. Similarly, most of the evidence sources to inform transition probabilities from the remaining HSCT health states are not described in the CS (neither are the transition probabilities). This includes assumptions regarding the extrapolation of treatment effectiveness which is not extensively discussed in the CS. The lack of detailed description and justification is worrying, given treatment effectiveness (including implicit assumptions made and selection of evidence sources to obtain transition probabilities) is often an influential part of cost effectiveness models.

In the company base-case (probabilistic) AATO is less expensive (£31,088 saved) and more effective (2.546 QALYs gained) than AIDA and thus the dominating strategy for newly diagnosed low-to-intermediate risk APL (i.e. the first line population). AATO remained dominant in most of the sensitivity and scenario analyses conducted by the company. The ERG has incorporated various adjustments to the company base-case this resulted in the (deterministic) ERG base-case wherein AATO remained dominant. Moreover, the ERG produced a worst-case scenario (combination of some of the scenario analyses explored by the ERG), to acknowledge the uncertainties discussed in section 5.2 of this report. This resulted in an ICER of £21,622 per QALY gained (deterministic). The ERG was unable to perform probabilistic analysis for its base-case. However, the ERG does not consider this to be a major issue as AATO is likely to remain dominant if the ERG would be able to produce probabilistic results for its base-case.

In conclusion, despite the ERG's criticism of the economic model and several highlighted uncertainties, it is reassuring that AATO for the first line population remained dominant in the ERG base-case, and that the worst-case scenario produced by the ERG resulted in an ICER of £21,622. However, as indicated by the subgroup analysis performed by the ERG, the cost effectiveness of AATO for the second line might be substantially different (estimated ICER of £31,184 per QALY gained).

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

In Section 5.3 the ERG base-case was presented, which was based on various changes compared to the company base-case. Table 6.1 shows how individual changes impact the results plus the combined effect of all changes simultaneously. The exploratory scenario analyses are presented in Table 6.2. These are all conditional on the ERG base-case. The analyses numbers in Tables 6.1 and 6.2 correspond to the analyses numbers reported in Section 5.3. Finally, Table 6.3 provides the results of the subgroup analysis (described in Section 5.3.3). The submitted model file contains technical details on the analyses performed by the ERG (e.g. the "ERG" sheet provides an overview of the cells that were altered for each adjustment).

Table 6.1: Deterministic ERG base-case

CS base-case					
Technologies	**Total costs**	**Total QALYs**	**Incremental costs**	**Incremental QALYs**	**ICER (£/QALY)**
AATO	£104,996	16.336			
AIDA	£136,267	13.717	-£31,270	2.618	Dominance
Fixing errors (1-6)					
AATO	£105,847	16.287			
AIDA	£131,760	13.859	-£25,914	2.428	Dominance
Extend time horizon (1-6, 7)[a]					
AATO	£106,722	16.777			
AIDA	£134,262	14.149	-£27,540	2.629	Dominance
Alternative utility values (1-6, 8)[a]					
AATO	£105,847	16.527			
AIDA	£131,760	14.116	-£25,914	2.411	Dominance
Capping utility values (1-6, 9)[a]					
AATO	£105,847	15.598			
AIDA	£131,760	13.338	-£25,914	2.260	Dominance
Remove inappropriate parameters from PSA (1-6, 10)[a]					
AATO	£105,847	16.287			
AIDA	£131,760	13.859	-£25,914	2.428	Dominance
Alternative remission probabilities (1-6, 11)[a]					
AATO	£106,055	16.280			
AIDA	£127,908	14.015	-£21,853	2.265	Dominance
ERG base-case (1-11)					
AATO	£106,931	16.135			
AIDA	£130,432	13.881	-£23,502	2.254	Dominance
[a]Analyses performed conditional on the fixing error analysis.					

Table 6.2: Deterministic scenario analyses conditional on ERG base-case

ERG base-case (1-11)					
Technologies	Total costs	Total QALYs	Incremental costs	Incremental QALYs	ICER (£/QALY)
AATO	£106,931	16.135			
AIDA	£130,432	13.881	-£23,502	2.254	Dominance
Disease-related mortality during the induction phase (1-11, 12)					
AATO	£103,532	15.530			
AIDA	£120,599	12.848	-£17,066	2.682	Dominance
Relapse probability equal for all treatments two year after first line remission (1-11, 13)					
AATO	£106,931	16.135			
AIDA	£86,524	15.100	£20,407	1.034	£19,734
Transitions to alloHSCT replaced for transitions to autoHSCT (1-11, 14)					
AATO	£103,523	16.283			
AIDA	£113,388	14.659	-£9,865	1.624	Dominance
Transitions from second line remission to HSCT states removed (1-11, 15)					
AATO	£107,200	16.129			
AIDA	£132,049	13.849	-£24,848	2.281	Dominance
Transitions from second line remission to HSCT states ‘uncorrected (1-11, 16)					
AATO	£106,773	16.137			
AIDA	£129,496	13.895	-£22,723	2.242	Dominance
Cardiac events added for AATO to reflecting treatment switching due to (potential) arrhytmia (1-11, 17)					
AATO	£107,285	16.121			
AIDA	£130,891	13.836	-£23,606	2.285	Dominance
Worst-case scenario (1-15, 17)					
AATO	£100,561	15.662			
AIDA	£73,494	14.410	£27,067	1.252	£21,622

Table 6.3: Deterministic subgroup analysis reflecting the second line population

Second line population (conditional on ERG base-case)					
Technologies	**Total costs**	**Total QALYs**	**Incremental costs**	**Incremental QALYs**	**ICER (£/QALY)**
AATO	£209,365	9.204			
AIDA	£191,158	8.620	£18,207	0.584	£31,184

7. OVERALL CONCLUSIONS

7.1 Statement of principal findings

The company presented evidence from three RCTs: Two of these were trials in newly diagnosed APL (APL0406 and AML17) and the third was a study in patients with relapsed APL (Raffoux, et al. 2003).

Untreated APL

Both trials in newly diagnosed APL (APL0406 and AML17) compared AATO (all-trans retinoic acid (ATRA) + ATO) with AIDA (ATRA + idarubicin). The dosing and regimens for AATO in AML17 were not in accordance with the licence. As the dosing and regimens for AATO in APL0406 were in accordance with the licence the ERG focused on this trial. APL0406 included 266 patients with newly-diagnosed, low- to intermediate-risk APL aged 18 to 71 years and took place in Italy and Germany.

Results from APL0406 showed that AATO significantly improved overall survival (OS) at 50 months compared with AIDA (99.2% vs 92.6% respectively, p=0.007). The primary endpoint of this trial was event-free survival (EFS) at two years in the initial cohort of 156 patients (97% with AATO vs 86% with AIDA, p<0.001 for non-inferiority, p=0.02 for superiority). EFS was significantly better in the AATO group across all subsequent analyses to reach 97.3% at 50 months in the full cohort of 266 patients, compared with 80.0% in the AIDA group (p<0.001). At 50 months, the cumulative incidence of relapse was 1.9% in the AATO group compared with 13.9% in the AIDA group (p=0.0013). In the AATO group patients experienced fewer haematological adverse events including fever and infection episodes and grade 3 to 4 neutropenia and thrombocytopenia lasting over 15 days. AATO was also more favourable than AIDA for grade 3-4 gastrointestinal toxicity. Other adverse events were more common with AATO mainly in the induction phase of treatment. In the AATO group incidence of leukocytosis was 43%. A greater number of patients experienced QTc prolongation with AATO (11% vs 0.7%). In addition, a greater number of patients experienced grade 3 to 4 hepatic toxicity, (40% vs. 3%). There were no instances of neurotoxicity with AIDA but 19 instances were noted with AATO.

Relapsed or refractory APL

The only trial presented for relapsed/refractory patients was by Raffoux et al. (2003). This small trial compared AATO with ATO, which is not a relevant comparison according to the scope. OS was similar between the AATO and ATO study arms. Across both groups, the estimated two-year OS was 59% (95% CI: 35%–77%). EFS was not reported in this study.

Economic evaluation

In the company base-case (probabilistic) AATO is less expensive (£31,088 saved) and more effective (2.546 QALYs gained) than AIDA and thus the dominating strategy for newly diagnosed low-to-intermediate risk APL (i.e. the first line population). AATO remained dominant in most of the sensitivity and scenario analyses conducted by the company. The ERG has incorporated various adjustments to the company base-case this resulted in the (deterministic) ERG base-case wherein AATO remained dominant. Moreover, the ERG produced a worst-case scenario (combination of some of the scenario analyses explored by the ERG), to acknowledge the uncertainties discussed in section 5.2 of this report. This resulted in an ICER of £21,622 per QALY gained (deterministic). The ERG was unable to perform probabilistic analysis for its base-case. However, the ERG does not consider this to be a major issue as AATO is likely to remain dominant if the ERG would be able to produce probabilistic results for its base-case.

In conclusion, despite the ERG's criticism of the economic model and several highlighted uncertainties, it is reassuring that AATO for the first line population remained dominant in the ERG base-case, and

that the worst-case scenario produced by the ERG resulted in an ICER of £21,622. However, as indicated by the subgroup analysis performed by the ERG, the cost effectiveness of AATO for the second line might be substantially different (estimated ICER of £31,184 per QALY gained).

7.2 Strengths and limitations of the assessment

Overall, the company submission searches were well presented and reproducible. Searches were carried out on a range of databases and supplementary resources. However, the ERG was concerned about the overall quality of the searches conducted, as there were numerous inconsistencies, inaccuracies and redundancy throughout. It is, thus, possible that relevant evidence may have been missed. However, the main weakness of the submission is that only one trial is directly relevant to the appraisal (APL0406) which provides data on an untreated population only. The trial does not have any UK patients. The company presented one trial in relapsed/refractory patients. However, the trial did not present a relevant comparison according to the NICE scope. The committee will need to consider whether it is necessary to explore further the evidence for relapsed/refractory patients or whether it is sufficiently well-established in routine clinical practice.

Strengths related to the economic evaluation include the granularity the model structure provides in comparison with other CEAs identified in the SLR. However, related to this, the (lack of) data to inform post first line transition probabilities can be regarded as a limitation. Additionally, the lack of (EQ5D) utility values for the APL population is a concern. Nevertheless, it is reassuring that AATO for the first line population remained dominant in the ERG base-case, and that the worst-case scenario produced by the ERG resulted in an (deterministic) ICER of £21,622 per QALY gained.

7.3 Suggested research priorities

Although decision uncertainty in the economic evaluation is relatively low, suggested research priorities regarding the cost effectiveness might be focused on obtaining health state utility values for the APL population as well as transition probabilities from and to the HSCT health states reflective of UK clinical practice.

8. REFERENCES

[1] Teva Pharma B.V. *Arsenic trioxide for treating acute promyelocytic leukaemia - TA10216. Document B. Company submission from Teva. Single technology appraisal (STA)*: Teva Pharma B.V., 2017

[2] National Institute for Health and Care Excellence. *Arsenic trioxide for treating acute promyelocytic leukaemia. Final scope [Internet]*. London: National Institute for Health and Care Excellence, 2017 [accessed 8.12.17] Available from: https://www.nice.org.uk/guidance/gid-ta10216/documents/final-scope

[3] Sant M, Allemani C, Tereanu C, De Angelis R, Capocaccia R, Visser O, et al. Incidence of hematologic malignancies in Europe by morphologic subtype: results of the HAEMACARE project. *Blood* 2010;116(19):3724-34.

[4] Visser O, Trama A, Maynadié M, Stiller C, Marcos-Gragera R, De Angelis R, et al. Incidence, survival and prevalence of myeloid malignancies in Europe. *Eur J Cancer* 2012;48(17):3257-66.

[5] Dores GM, Devesa SS, Curtis RE, Linet MS, Morton LM. Acute leukemia incidence and patient survival among children and adults in the United States, 2001-2007. *Blood* 2012;119(1):34-43.

[6] Rashidi A, Riley M, Goldin TA, Sayedian F, Bayerl MG, Aguilera NS, et al. Delay in the administration of all-trans retinoic acid and its effects on early mortality in acute promyelocytic leukemia: final results of a multicentric study in the United States. *Leuk Res* 2014;38(9):1036-40.

[7] Sanz MA, Lo Coco F, Martín G, Avvisati G, Rayón C, Barbui T, et al. Definition of relapse risk and role of nonanthracycline drugs for consolidation in patients with acute promyelocytic leukemia: a joint study of the PETHEMA and GIMEMA cooperative groups. *Blood* 2000;96(4):1247-53.

[8] Fey MF, Buske C, ESMO Guidelines Working Group. Acute myeloblastic leukaemias in adult patients: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol* 2013;24 Suppl 6:vi138-43.

[9] Sanz MA, Grimwade D, Tallman MS, Lowenberg B, Fenaux P, Estey EH, et al. Management of acute promyelocytic leukemia: recommendations from an expert panel on behalf of the European LeukemiaNet. *Blood* 2009;113(9):1875-91.

[10] Lengfelder E, Lambert J-F, Nachbaur D, Niederwieser D, Platzbecker U, Schlenk RF, et al. *Onkopedia. Akute promyelozyten leukämie (APL) [Internet]*. Berlin: DGHO Deutsche Gesellschaft für Hämatologie und Medizinische Onkologie e.V., 2014 [accessed 15.12.17] Available from: https://www.onkopedia.com/de/onkopedia/guidelines/akute-promyelozyten-leukaemie-apl/@@view/html/index.html

[11] Kompetenz bei Akuten und Chronischen Leukämien. *German Intergroup recommendations on the diagnostic and therapeutic management of acute promyelocytic leukemia (APL). Version: Final 3.0 [Internet]*. Mannheim: Kompetenz bei Akuten und Chronischen Leukämien, 2015 [accessed 15.12.17] Available from: https://www.kompetenznetz-leukaemie.de/sites/kompetenznetz-leukaemie/content/e53457/e55758/e57007/NapoleonRegister_TreatmentRecommendation_Final_V3.0_20150630.pdf

[12] Lo-Coco F, Avvisati G, Vignetti M, Thiede C, Orlando SM, Iacobelli S, et al. Retinoic acid and arsenic trioxide for acute promyelocytic leukemia. *N Engl J Med* 2013;369(2):111-21.

[13] Burnett AK, Russell NH, Hills RK, Bowen D, Kell J, Knapper S, et al. Arsenic trioxide and all-trans retinoic acid treatment for acute promyelocytic leukaemia in all risk groups (AML17): results of a randomised, controlled, phase 3 trial. *Lancet Oncol* 2015;16(13):1295-305.

[14] Platzbecker U, Avvisati G, Cicconi L, Thiede C, Paoloni F, Vignetti M, et al. Improved outcomes with retinoic acid and arsenic trioxide compared with retinoic acid and chemotherapy in non-high-risk

acute promyelocytic leukemia: final results of the randomized Italian-German APL0406 trial. *J Clin Oncol* 2017;35(6):605-12.

[15] Russell NH, Burnett AK, Hills RK, Betteridge S, Dennis M, Dillon R, et al. Long term follow up from the NCRI AML17 trial of attenuated arsenic trioxide and ATRA therapy for newly diagnosed and relapsed acute promyelocytic leukaemia. *Blood* 2016;128(22):897.

[16] Lengfelder E, European LeukemiaNet. *European recommendation for salvage therapy of relapsed acute promyelocytic leukemia (APL) including arsenic trioxide (ATO) [Internet]*: European LeukemiaNet, 2007 [accessed 15.12.17] Available from: https://www.leukemia-net.org/content/leukemias/aml/apl/apl_recommendations/e6479/infoboxContent6480/APLrelapserecommendation121207.pdf.

[17] Lengfelder E, Lo-Coco F, Ades L, Montesinos P, Grimwade D, Kishore B, et al. Arsenic trioxide-based therapy of relapsed acute promyelocytic leukemia: registry results from the European LeukemiaNet. *Leukemia* 2015;29(5):1084-91.

[18] Raffoux E, Rousselot P, Poupon J, Daniel MT, Cassinat B, Delarue R, et al. Combined treatment with arsenic trioxide and all-trans-retinoic acid in patients with relapsed acute promyelocytic leukemia. *J Clin Oncol* 2003;21(12):2326-34.

[19] Teva Pharma B.V. *Arsenic trioxide for treating acute promyelocytic leukaemia [ID446]. Response to request for clarification from the ERG*: Teva Pharma B.V., 2018

[20] European Medicines Agency. *Assessment report - variation. Trisenox. International non-proprietary name: arsenic trioxide. Procedure no. EMEA/H/C/000388/II/0058 [Internet]*. London: European Medicines Agency, 2016 [accessed 8.1.18] Available from: http://www.ema.europa.eu/docs/en_GB/document_library/EPAR_-_Assessment_Report_-_Variation/human/000388/WC500220071.pdf

[21] U.S. National Library of Medicine. Genetics Home Reference. Acute promyelocytic leukemia [Internet]. U.S. National Library of Medicine, [accessed 15.12.17]. Available from: https://ghr.nlm.nih.gov/condition/acute-promyelocytic-leukemia

[22] Higgins JPT, Green S, eds. *Cochrane handbook for systematic reviews of interventions [Internet]*. Version 5.1.0 [updated March 2011]: The Cochrane Collaboration, 2011 [accessed 8.1.18]. Available from: http://handbook.cochrane.org/

[23] Centre for Reviews and Dissemination. *Systematic Reviews: CRD's guidance for undertaking reviews in health care [Internet]*. York: University of York, 2009 [accessed 8.1.18] Available from: http://www.york.ac.uk/inst/crd/SysRev/!SSL!/WebHelp/SysRev3.htm

[24] Thomas X, Pigneux A, Raffoux E, Huguet F, Caillot D, Fenaux P. Superiority of an arsenic trioxide-based regimen over a historic control combining all-trans retinoic acid plus intensive chemotherapy in the treatment of relapsed acute promyelocytic leukemia. *Haematologica* 2006;91(7):996-7.

[25] Yanada M, Yano S, Kanamori H, Gotoh M, Emi N, Watakabe K, et al. Autologous hematopoietic cell transplantation for acute promyelocytic leukemia in second complete remission: outcomes before and after the introduction of arsenic trioxide. *Leuk Lymphoma* 2017;58(5):1061-7.

[26] Lo-Coco F, Avvisati G, Vignetti M, Breccia M, Gallo E, Rambaldi A, et al. Front-line treatment of acute promyelocytic leukemia with AIDA induction followed by risk-adapted consolidation for adults younger than 61 years: results of the AIDA-2000 trial of the GIMEMA Group. *Blood* 2010;116(17):3171-9.

[27] Ravandi F, Estey E, Jones D, Faderl S, O'Brien S, Fiorentino J, et al. Effective treatment of acute promyelocytic leukemia with all-trans-retinoic acid, arsenic trioxide, and gemtuzumab ozogamicin. *J Clin Oncol* 2009;27(4):504-10.

[28] Efficace F, Mandelli F, Avvisati G, Cottone F, Ferrara F, Di Bona E, et al. Randomized phase III trial of retinoic acid and arsenic trioxide versus retinoic acid and chemotherapy in patients with acute promyelocytic leukemia: health-related quality-of-life outcomes. *J Clin Oncol* 2014;32(30):3406-12.

[29] Burnett AK, Hills RK, Grimwade D, Jovanovic JV, Craig J, McMullin MF, et al. Inclusion of chemotherapy in addition to anthracycline in the treatment of acute promyelocytic leukaemia does not improve outcomes: results of the MRC AML15 trial. *Leukemia* 2013;27(4):843-51.

[30] US Food and Drug Administration. Search orphan drug designations and approvals. Arsenic trioxide [Internet]. Silver Spring, MD: US Food and Drug Administration, [accessed 15.12.17]. Available from: https://www.accessdata.fda.gov/scripts/opdlisting/oopd/detailedIndex.cfm?cfgridkey=109697

[31] Tallman M, Lo-Coco F, Barnes G, Kruse M, Wildner R, Martin M, et al. Cost-effectiveness analysis of treating acute promyelocytic leukemia patients with arsenic trioxide and retinoic acid in the United States. *Clin Lymphoma Myeloma Leuk* 2015;15(12):771-7.

[32] Russell N, Hills R, Betteridge S, Dennis M, Dillon R, Grimwade D, et al. *The feasibility of more attenuated dosage schedules of ATO in newly diagnosed and relapsed acute promyelocytic leukaemia. Data on file*, 2017

[33] Hosing C, Saliba RM, McLaughlin P, Andersson B, Rodriguez MA, Fayad L, et al. Long-term results favor allogeneic over autologous hematopoietic stem cell transplantation in patients with refractory or recurrent indolent non-Hodgkin's lymphoma. *Ann Oncol* 2003;14(5):737-44.

[34] Ramadan SM, Di Veroli A, Camboni A, Breccia M, Iori AP, Aversa F, et al. Allogeneic stem cell transplantation for advanced acute promyelocytic leukemia in the ATRA and ATO era. *Haematologica* 2012;97(11):1731-5.

[35] de Botton S, Fawaz A, Chevret S, Dombret H, Thomas X, Sanz M, et al. Autologous and allogeneic stem-cell transplantation as salvage treatment of acute promyelocytic leukemia initially treated with all-trans-retinoic acid: a retrospective analysis of the European Acute Promyelocytic Leukemia Group. *J Clin Oncol* 2005;23(1):120-6.

[36] Holter Chakrabarty JL, Rubinger M, Le-Rademacher J, Wang HL, Grigg A, Selby GB, et al. Autologous is superior to allogeneic hematopoietic cell transplantation for acute promyelocytic leukemia in second complete remission. *Biol Blood Marrow Transplant* 2014;20(7):1021-5.

[37] Ma X, Does M, Raza A, Mayne ST. Myelodysplastic syndromes: incidence and survival in the United States. *Cancer* 2007;109(8):1536-42.

[38] Lachaine J, Mathurin K, Barakat S, Schuh AC. Economic evaluation of arsenic trioxide for treatment of newly diagnosed acute promyelocytic leukaemia in Canada. *Hematol Oncol* 2015;33(4):229-38.

[39] Douer D, Tallman MS. Arsenic trioxide: new clinical experience with an old medication in hematologic malignancies. *J Clin Oncol* 2005;23(10):2396-410.

[40] Wolff SN, Herzig RH, Fay JW, Phillips GL, Lazarus HM, Flexner JM, et al. High-dose cytarabine and daunorubicin as consolidation therapy for acute myeloid leukemia in first remission: long-term follow-up and results. *J Clin Oncol* 1989;7(9):1260-7.

[41] Fenaux P, Le Deley MC, Castaigne S, Archimbaud E, Chomienne C, Link H, et al. Effect of all transretinoic acid in newly diagnosed acute promyelocytic leukemia. Results of a multicenter randomized trial. European APL 91 Group. *Blood* 1993;82(11):3241-9.

[42] WebMD. Pneumonia - What happens [Internet]. [accessed 29.1.18]. Available from: http://www.webmd.com/lung/tc/pneumonia-what-happens

[43] Shoenfeld Y, Gurewich Y, Gallant LA, Pinkhas J. Prednisone-induced leukocytosis. Influence of dosage, method and duration of administration on the degree of leukocytosis. *Am J Med* 1981;71(5):773-8.

[44] Zhu HH, Wu DP, Jin J, Li JY, Ma J, Wang JX, et al. Oral tetra-arsenic tetra-sulfide formula versus intravenous arsenic trioxide as first-line treatment of acute promyelocytic leukemia: a multicenter randomized controlled trial. *J Clin Oncol* 2013;31(33):4215-21.

[45] Ratnaike RN. Acute and chronic arsenic toxicity. *Postgrad Med J* 2003;79(933):391-6.

[46] Breccia M, Latagliata R, Carmosino I, Cannella L, Diverio D, Guarini A, et al. Clinical and biological features of acute promyelocytic leukemia patients developing retinoic acid syndrome during induction treatment with all-trans retinoic acid and idarubicin. *Haematologica* 2008;93(12):1918-20.

[47] Mathews VB, P., Shaji RV, George B, Chandy M, Srivastava A. Arsenic trioxide in the treatment of newly diagnosed acute promyelocytic leukemia: a single center experience. *Am J Hematol* 2002;70(4):292-9.

[48] Siu CW, Au WY, Yung C, Kumana CR, Lau CP, Kwong YL, et al. Effects of oral arsenic trioxide therapy on QT intervals in patients with acute promyelocytic leukemia: implications for long-term cardiac safety. *Blood* 2006;108(11):103-6.

[49] Woods B, Hawkins N, Dunlop W, O'Toole A, Bramham-Jones S. Bendamustine versus chlorambucil for the first-line treatment of chronic lymphocytic leukemia in England and Wales: a cost-utility analysis. *Value Health* 2012;15(5):759-70.

[50] Beusterien KM, Davies J, Leach M, Meiklejohn D, Grinspan JL, O'Toole A, et al. Population preference values for treatment outcomes in chronic lymphocytic leukaemia: a cross-sectional utility study. *Health Qual Life Outcomes* 2010;8:50.

[51] Szende A, Janssen B, Cabases J, eds. *Self-reported population health: an international perspective based on EQ-5D [Internet]*. Dordrecht: SpringerOpen, 2014 [accessed 20.12.17]. Available from: https://link.springer.com/book/10.1007/978-94-007-7596-1

[52] Breitscheidel L. Cost utility of allogeneic stem cell transplantation with matched unrelated donor versus treatment with imatinib for adult patients with newly diagnosed chronic myeloid leukaemia. *J Med Econ* 2008;11(4):571-84.

[53] Morton RL, Howard K, Webster AC, Wong G, Craig JC. The cost-effectiveness of induction immunosuppression in kidney transplantation. *Nephrol Dial Transplant* 2009;24(7):2258-69.

[54] Cooperberg MR, Ramakrishna NR, Duff SB, Hughes KE, Sadownik S, Smith JA, et al. Primary treatments for clinically localised prostate cancer: a comprehensive lifetime cost-utility analysis. *BJU Int* 2013;111(3):437-50.

[55] Attard CL, Brown S, Alloul K, Moore MJ. Cost-effectiveness of folfirinox for first-line treatment of metastatic pancreatic cancer. *Curr Oncol* 2014;21(1):e41-51.

[56] Stevenson SM, Danzig MR, Ghandour RA, Deibert CM, Decastro GJ, Benson MC, et al. Cost-effectiveness of neoadjuvant chemotherapy before radical cystectomy for muscle-invasive bladder cancer. *Urol Oncol* 2014;32(8):1172-7.

[57] Choi HW, Miele K, Dowdy D, Shah M. Cost-effectiveness of Xpert(R) MTB/RIF for diagnosing pulmonary tuberculosis in the United States. *Int J Tuberc Lung Dis* 2013;17(10):1328-35.

[58] Prica A, Chan K, Cheung M. Combined modality therapy versus chemotherapy alone as an induction regimen for primary central nervous system lymphoma: a cost-effectiveness analysis. *Neuro Oncol* 2014;16(10):1384-91.

[59] Nshimyumukiza L, Duplantie J, Gagnon M, Douville X, Fournier D, Lindsay C, et al. Dabigatran versus warfarin under standard or pharmacogenetic-guided management for the prevention of stroke and systemic thromboembolism in patients with atrial fibrillation: a cost/utility analysis using an analytic decision model. *Thrombosis Journal* 2013;11(1):14.

[60] National Institute for Health and Care Excellence. *Arsenic trioxide for treating acute promyelocytic leukaemia [ID446]. Clarification letter*. London: National Institute for Health and Care Excellence, 2018

[61] Joint Formulary Committee. *British National Formulary. 73*. London: BMJ Group and Pharmaceutical Press, 2017.

[62] Department of Health. *NHS reference costs 2014-2015 [Internet]*. London: Department of Health, 2015 [accessed 20.12.17] Available from: https://www.gov.uk/government/publications/nhs-reference-costs-2014-to-2015

[63] Morgan A, Sutton A, Wailoo A. *The risk and costs of febrile neutropenia in patients with non-small cell lung cancer treated with docetaxel. Report by the NICE Decision Support Unit [Internet]*. Sheffield: NICE Decision Support Unit, 2007 [accessed 20.12.17] Available from: http://nicedsu.org.uk/

[64] Soini E, Hautala A, Poikonen E, Becker U, Kyttälä M, Martikainen J. Cost-effectiveness of first-line chronic lymphocytic leukemia treatments when full-dose fludarabine is unsuitable. *Clin Ther* 2016;38(4):889-904.e14.

[65] Akhtar W, Chung Y. Saving the NHS one blood test at a time. *BMJ Qual Improv Rep* 2014;2(2):pii: u204012.w1749.

[66] Calhoun EA, Chang CH, Welshman EE, Fishman DA, Lurain JR, Bennett CL. Evaluating the total costs of chemotherapy-induced toxicity: results from a pilot study with ovarian cancer patients. *Oncologist* 2001;6(5):441-5.

[67] Milligan DW, Grimwade D, Cullis JO, Bond L, Swirsky D, Craddock C, et al. Guidelines on the management of acute myeloid leukaemia in adults. *Br J Haematol* 2006;135(4):450-74.

[68] National Institute for Health and Care Excellence. *Chronic heart failure: costing report*. London: National Institute for Health and Care Excellence, 2010

[69] National Institute for Health and Care Excellence. *Intravenous fluid therapy in adults in hospital. NICE Clinical guideline 174 [Internet]*. London: National Institute for Health and Care Excellence, 2013 [accessed 20.12.17] Available from: https://www.nice.org.uk/guidance/cg174

[70] Saito AM, Cutler C, Zahrieh D, Soiffer RJ, Ho VT, Alyea EP, et al. Costs of allogeneic hematopoietic cell transplantation with high-dose regimens. *Biol Blood Marrow Transplant* 2008;14(2):197-207.

[71] Jones CA, Fernandez LP, Weimersheimer P, Zakai NA, Sharf M, Mesa OA, et al. Estimating the burden of cost in chronic graft-versus-host disease: a human capital approach. *JHEOR* 2016;4(2):113-8.

[72] Office for National Statistics (ONS). *National Life Tables, England & Wales, 1980-82 to 2014-16*: Office for National Statistics

[73] Kind P, Hardman G, Macran S. *UK population norms for EQ-5D. Discussion paper 172 [Internet]*. York: Centre for Health Economics, University of York, 1999 [accessed 15.2.18] Available from:

https://www.york.ac.uk/media/che/documents/papers/discussionpapers/CHE%20Discussion%20Paper%20172.pdf

[74] Kaltenthaler E, Carroll C, Hill-McManus D, Scope A, Holmes M, Rice S, et al. The use of exploratory analyses within the National Institute for Health and Care Excellence single technology appraisal process: an evaluation and qualitative analysis. *Health Technol Assess* 2016;20(26):1-48.

Appendix 1: ERG search strategies

Detailed critique of clinical effectiveness searches:

- The database searches were clearly structured (population and study design), using a combination of subject heading indexing and free text terms, with synonyms, adjacency operators and truncation.
- The population facet would have been improved by introducing a specified number to the adjacency operator. When using *adj* without a number affixed the search terms must appear next to each other in that order; affixing a number finds the search terms in any order, within the specified number of words. For example, using ***adj3*** in the search line '(promyelocyt* adj (leukaemia or leukemia)).mp.' would have increased sensitivity by identifying records with 'promyelocytic acute leukaemia' and 'leukemia, acute promyelocytic'.
- The search terms for leukaemia could have been truncated to increase sensitivity, e.g. leukaemia$ or leukemia$
- Additional synonyms and acronyms could have been included in the search strategies, e.g. APML, APL, AML M3, ANLL M3, progranulocytic leukaemia.
- The full date span for the databases searched was not provided.
- The eligibility criteria provided in Table 2.1 of the company submission included systematic reviews and meta-analyses, but no attempt to search for these study designs was made. Search terms for systematic reviews were not included in the strategies, and systematic review specific resources such as the Cochrane Database of Systematic Reviews (CDSR) and the Database of Abstracts of Reviews of Effects (DARE) were not searched.
- The search strategy provided in Appendix D reported a simultaneous search of MEDLINE and Embase using the Ovid interface. A simultaneous multi-file search such as this should include both MeSH and EMTREE subject headings to ensure that all subject indexing terms are searched; the search strategy only included the EMTREE term 'promyelocytic leukemia/' for the initial search, then only the MeSH term 'exp Leukemia, Promyelocytic, Acute/' for the update search. In this case, the EMTREE term does not map to the equivalent MeSH term when conducting a simultaneous multi-file search, whereas the MeSH term does map to the EMTREE term. Although the EMTREE term 'promyelocytic leukemia/' is reported in Appendix D, Table 1.1, the results would indicate that the MeSH term was actually used in the search. Indeed, when a simultaneous search of MEDLINE and Embase is conducted in Ovid the following message appears: *[Ovid MEDLINE] – The subject heading 'promyelocytic leukemia' is invalid in this database.*
- It appears that the RCT filter used was based on the Cochrane Highly Sensitive Search Strategy for identifying randomized trials in MEDLINE; this was not explicitly stated in either the clinical effectiveness section of the company submission (B.2) or in Appendix D. This search filter was designed specifically for use in MEDLINE, and does not translate to work efficiently in Embase.
- The company submission attempted to identify safety data as well as clinical effectiveness data by conducting two separate literature searches. The study design terms used to search for non-RCTs were possibly too restrictive to capture all safety data. It is not clear where the search terms used to search for non-RCTs were derived from. EMTREE subject heading terms were included, but not MeSH terms. Although the strategy was not limited to RCTs, it was still limited to study designs that do not necessarily capture safety data (longitudinal studies, retrospective studies, prospective studies, follow-up studies). CRD guidance[23] recommends that if searches have been limited by a study design filter, additional searches should be undertaken to ensure that adverse events that are long-term, rare or unanticipated are not missed. Ideally,

this would entail searching without any study design terms, or would include generic and specific adverse event and safety search terms.

- There were a number of redundant lines included in the search strategies, e.g. Line 3 in Table 1.1, Lines 3 and 4 in Table 1.2 (Appendix D)
- The method used to limit the MEDLINE and Embase searches to human studies was incorrect. The strategy included the line '*Animals.sh.*' and then used the Boolean operator *NOT* to remove the records identified. The correct limit should be '*exp animals/ not humans.sh.*' when searching MEDLINE, or for a simultaneous multi-file search the automatic limit provided by Ovid should be used: *Human*. It is possible that potentially relevant studies were excluded from the final search results using this approach, as records including terms for both human and animal would have been omitted.
- The MEDLINE and Embase search strategy used a variety of different field tags (mp, tw, ti, ab) when a more consistent approach is used in current best practice.
- It is not clear which database was searched in the Cochrane Library for RCTs. CENTRAL should have been searched for RCTs, but from the results reported in the strategy it would appear that CDSR was searched instead. In response to the ERG clarification letter, the company confirmed that CENTRAL was searched (though the results reported in Table 1.3 of Appendix D would suggest otherwise).
- For the searches of conference proceedings the company submission did not provide full details of the search terms used, the precise date of the searches or the number of records retrieved. Details were provided by the company in response to the ERG clarification letter.
- The October 2017 update searches did not include the specific date ranges searched in MEDLINE and Embase.
- The date limit used for the MEDLINE and Embase update searches was unusual: a line for publication date 1968 to 2015 was combined with Boolean *NOT* to identify studies published from 2016 to 2017, when simply limiting to 2016 to 2017 would have been sufficient.
- The results reported in the search strategies did not correspond with those presented in the PRISMA flow charts: Figure 2.1 and Figure 2.2 in section B.2.

Detailed critique of cost effectiveness searches:

- The database searches were clearly structured (population and study design), using a combination of subject heading indexing and free text terms, with synonyms, adjacency operators and truncation.
- As per the clinical effectiveness search comments above (4.1.1), better use of adjacency, truncation and synonyms would have increased the sensitivity of the searches.
- The full date span for the databases searched was not provided.
- The search strategy reported a simultaneous search of MEDLINE and Embase using the Ovid interface. A simultaneous multi-file search such as this should ideally include both MeSH and EMTREE subject headings. EMTREE subject heading terms were included in the population facet, but not MeSH terms; whereas MeSH terms were included in the cost effectiveness facet, but not EMTREE. As with the clinical effectiveness searches, the EMTREE term '*exp promyelocytic leukemia/*' was reported in the initial 2016 search (Appendix G, Table 1.1), whilst the MeSH term '*exp Leukemia, Promyelocytic, Acute/'* was reported in the update search of October 2017 (Appendix G, Table 1.3).
- It is not clear where the search terms used for the cost effectiveness facet were derived from.
- The method used to limit the MEDLINE and Embase search to human studies was incorrect. The strategy included the line '*Animals.sh.*' and then used the Boolean operator *NOT* to remove

the records identified. The correct limit should be *'exp animals/ not humans.sh.'* when searching MEDLINE. The automatic Ovid limit '*Human'* would have been a better option for this simultaneous multi-file search. It is possible that potentially relevant studies were excluded from the final search results using this approach, as records using terms for both human and animal would have been omitted.

- The MEDLINE and Embase search strategy used a variety of different field tags (mp, tw, ti, ab) when a more consistent approach is used in current practice.
- The host provider used to search NHS EED was not reported. The company responded to the ERG clarification letter to confirm that the CRD interface was used to search NHS EED.
- There was no reason to conduct an update search of NHS EED in October 2017, as this database ceased in April 2015 (Issue 2 of 4). The initial search was conducted in July 2016.
- The date limit used for the MEDLINE and Embase update searches was unusual: a line for publication date 1968 to 2015 was combined with Boolean *NOT* to identify studies published from 2016 to 2017. Limiting to 2016 to 2017 would have been sufficient.
- A search of health economic databases, such as the Cost Effectiveness Analysis (CEA) Registry (www.cearegistry.org) and ScHARRHUD (http://www.scharrhud.org/), would have been a useful addition to the literature searches.

Detailed critique of measurement and valuation of health effects searches:

- The database searches were clearly structured (population and study design), using a combination of subject heading indexing and free text terms, with synonyms, adjacency operators and truncation.
- Again, as with the searches conducted for clinical effectiveness and cost effectiveness, better use of adjacency, truncation and synonyms could have been made to increase the sensitivity of the search results.
- The full date span for the databases searched was not provided.
- The search strategy reported a simultaneous search of MEDLINE and Embase using the Ovid interface. A simultaneous multi-file search such as this should include both MeSH and EMTREE subject headings. EMTREE subject heading terms were included in the population facet, but not MeSH terms; whereas MeSH terms were included in the health-related quality-of-life and utilities facet, but no EMTREE terms were included.
- It is not clear where the search terms used for the health-related quality-of-life and utilities facet were derived from.
- There were a number of redundant lines included in the search strategies, e.g. Lines 5, 25, 26, 28 and 36 in Table 1.1, Lines 1 and 3 in Table 1.2 (Appendix H).
- The method used to limit the MEDLINE and Embase search to human studies was incorrect. The strategy included the line '*Animals.sh.*' and then used the Boolean operator *NOT* to remove the records identified. The correct limit should be '*exp animals/ not humans.sh.*' for searching in MEDLINE. The automatic Ovid limit '*Human'* would have been a better option for this simultaneous multi-file search. It is possible that potentially relevant studies were excluded from the final search results using this approach, as records using terms for both human and animal would have been omitted.
- The host provider used to search NHS EED was not reported. The company responded to the ERG clarification letter to confirm that the CRD interface was used to search NHS EED. A mixture of both CRD and Ovid search syntax was reported in the strategy.
- The NHS EED strategy included a facet of search terms for health-related quality-of-life and utilities; restricting the search unnecessarily.

- There was no reason to conduct an update search of NHS EED, as this database ceased in April 2015 (Issue 2 or 4), and the initial search was conducted in July 2016.
- The date limit used for the MEDLINE and Embase update searches was unusual: a line for publication date 1968 to 2015 was combined with Boolean *NOT* to identify studies published from 2016 to 2017. Limiting to 2016 to 2017 would have been sufficient.
- A search of health economic databases, such as Cost Effectiveness Analysis (CEA) Registry (www.cearegistry.org) and ScHARRHUD (http://www.scharrhud.org/), for utilities data would have been a useful addition to the literature searches

Detailed critique of cost and healthcare resource identification searches:

- The database searches were clearly structured (population and study design), using a combination of subject heading indexing and free text terms, with synonyms, adjacency operators and truncation.
- The full date span for the databases searched was not provided.
- The search strategy reported a simultaneous search of MEDLINE and Embase using the Ovid interface. A simultaneous multi-file search such as this should include both MeSH and EMTREE subject headings; the strategy only included the EMTREE term '*promyelocytic leukemia/'* in the initial 2016 search, whilst only including the MeSH term '*exp Leukemia, Promyelocytic, Acute/*' in the October 2017 update search.
- It is not clear where the search terms used for the resource use and costs facet were derived from.
- The method used to limit the MEDLINE and Embase search to human studies was incorrect. The strategy included the line '*Animals.sh*.' and then used the Boolean operator *NOT* to remove the records identified. The correct limit should be '*exp animals/ not humans.sh.*' when searching MEDLINE, whilst the automatic Ovid limit '*Human'* would have been preferable for a simultaneous multi-file search such as this. It is possible that potentially relevant studies were excluded from the final search results using this approach, as records using terms for both human and animal would have been omitted.
- The host provider used to search NHS EED was not reported. The company responded to the ERG clarification letter to confirm that the CRD interface was used to search NHS EED.
- The update search of NHS EED was unnecessary as this database ceased in April 2015 (Issue 2 of 4).
- The date limit used for the MEDLINE and Embase update searches was unusual: a line for publication date 1968 to 2015 was combined with Boolean *NOT* to identify studies published from 2016 to 2017. Limiting to 2016 to 2017 would have been sufficient.