Arsenic Trioxide for Treating Acute Promyelocytic Leukaemia: An Evidence Review Group Perspective of a NICE Single Technology Appraisal

Bram L. T. Ramaekers 1, Rob Riemsma 2, Sabine Grimm 1, Debra Fayter 2, Sohan Deshpande 2, Nigel Armstrong 2, Willem Witlox 1, Xavier Pouwels 1, Steven Duffy 2, Gill Worthy 2, Jos Kleijnen 2,3, Manuela A. Joore 1,3

Abstract

The National Institute for Health and Care Excellence (NICE) invited Teva, the company manufacturing arsenic trioxide (ATO; tradename Trisenox®), to submit evidence for the clinical and cost effectiveness of ATO for untreated and relapsed or refractory acute promyelocytic leukaemia (APL). Kleijnen Systematic Reviews Ltd (KSR), in collaboration with Maastricht University Medical Center, was commissioned as the independent Evidence Review Group (ERG). This paper presents a summary of the company submission (CS), the ERG’s critical review of the clinical and cost effectiveness evidence in the CS, key methodological considerations and the development of the NICE guidance by the Appraisal Committee (AC). The CS presented three randomized controlled trials (RCTs). Two of these were trials in newly diagnosed APL (APL0406 and AML17) and the third trial was in patients with relapsed APL. Results from APL0406 showed that more people having AATO [ATO plus all-trans retinoic acid (ATRA)] were alive at 50 months compared with people having AIDA (ATRA in combination with idarubicin) (99% vs. 93%; p = 0.007). There was also a statistically significant lower cumulative incidence of relapse with AATO compared with AIDA at 50 months (2% vs. 14%; p = 0.001). At 4 years, results from AML17 showed a significant difference in event-free survival (91% vs. 70%; p = 0.002) favouring AATO but not in overall survival (93% vs. 89%; p = 0.250). The only trial presented for relapsed/refractory patients compared AATO with ATO, which was not a relevant comparison according to the NICE scope. The AC concluded that AATO was effective for untreated APL while for relapsed or refractory APL the effectiveness of ATO was considered uncertain and the long-term safety remains unexplored. In the CS base-case, AATO was less expensive (£31,088 saved) and more effective (2.546 quality-adjusted life-years (QALYs) gained) than AIDA and thus the dominating strategy for newly diagnosed low- to intermediate-risk APL. However, the ERG’s critical assessment highlighted a number of concerns, including deviations from the NICE reference case and a lack of detailed description and justification of parameters and assumptions related to (the extrapolation of) treatment effectiveness. However, it was reassuring that AATO for untreated APL remained dominant in the ERG base-case, and that the worst-case scenario produced by the ERG resulted in an incremental cost-effectiveness ratio (ICER) of £21,622. The AC concluded that although there was uncertainty in the model, it could recommend ATO for both untreated and relapsed or refractory APL.
Key Points for Decision Makers

Arsenic trioxide plus all-trans retinoic acid was effective in terms of event-free survival for untreated acute promyelocytic leukaemia (APL), based on the two trials APL0406 and AML17, but the long-term safety of arsenic trioxide remains to be explored.

Although, in decision analytic models, uncertainty is ideally parameterised and considered in probabilistic sensitivity analyses, scenario analyses are informative to consider uncertainty due to indirectness or lack of evidence.

Age and sex-matched general population life-years (LYs) and quality-adjusted life-years (QALYs) can serve as a general validation benchmark for the LYs and QALYs estimated in decision analytic models.

The National Institute for Health and Care Excellence (NICE) Appraisal Committee recommended arsenic trioxide for both untreated and relapsed or refractory APL.

1 Introduction

Arsenic trioxide (ATO), tradename Trisenox®, was appraised within the National Institute for Health and Care Excellence (NICE) single technology appraisal process. Health technologies must be considered clinically effective and represent cost-effective use of National Health Service (NHS) resources in order to be recommended by NICE. As part of this, the company (Teva) provided NICE with a written submission, including an executable health economic model, detailing the company’s estimates of the clinical effectiveness and cost effectiveness of ATO for adults with acute promyelocytic leukaemia (APL). The evidence submitted by the company [1] was critically reviewed by an independent external organisation (Kleijnen Systematic Reviews) in collaboration with Maastricht University Medical Center, named the Evidence Review Group (ERG), which produced an ERG report [2]. After consideration of the submitted evidence, the NICE Appraisal Committee (AC) issued the Final Appraisal Determination, consisting of guidance regarding whether or not to recommend the technology, which is open to appeal [3]. This paper presents a summary of the ERG report and the development of the NICE guidance. It highlights key methodological issues emerging from the appraisal.

2 The Decision Problem

The NICE scope described the appraisal objective as ‘to appraise the clinical and cost effectiveness of ATO [with or without all-trans retinoic acid (ATRA)] within its marketing authorisation for adults with APL. More specifically, the scope defined two populations: (1) untreated low- to intermediate-risk APL; and (2) relapsed/refractory APL [4]. The population and intervention defined in the company submission (CS) were in line with the scope.

APL, a subtype of acute myeloid leukaemia, is a rare disease in which the production of blood cells in the bone marrow results in the development of abnormal white blood cells of the neutrophil type (promyelocytes). This leads to a reduction of normal white blood cells in circulation and eventually prevents the production of normal red blood cells and platelets. Initial symptoms of APL might appear inconspicuous, but the disease can progress rapidly with very poor survival. Long-term survival in APL remission can be achieved in ≥ 70% of newly diagnosed patients following chemotherapy-based treatment with the intention to achieve and maintain molecular remission [1].

The comparators listed in the NICE scope were:

1. AIDA regimen (ATRA in combination with idarubicin);
2. Haematopoietic stem-cell transplantation (HSCT) (for people with relapsed or refractory APL);
3. Best supportive care (BSC, for people with relapsed or refractory APL).

Consistent with only one population and one comparator in the scope, the CS assessed the cost effectiveness of ATRA+ATO (AATO) compared with AIDA in the newly diagnosed low- to intermediate-risk APL (i.e. first-line) population. The cost effectiveness of ATO in the relapsed/refractory APL (i.e. second-line) population was not assessed in the CS [1].

3 Independent Evidence Review Group (ERG) Review

Sections 3.1–3.4 summarise the evidence presented in the CS, as well as the ERG’s review of that evidence.

3.1 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) [5, 6] and the third was in patients with relapsed APL [7].

### 3.1.1 Newly Diagnosed Acute Promyelocytic Leukaemia (APL)

Both trials in newly diagnosed APL (APL0406 and AML17) compared AATO with AIDA. APL0406 included 266 patients with low- to intermediate-risk APL aged 18–71 years; while AML17 included 235 APL patients of any risk group, aged 16 years 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, event-free survival (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 3–4 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 3–4 neutropenia and thrombocytopenia lasting over 15 days.

Results from AML17 showed an EFS benefit of AATO over AIDA (4-year EFS of 91% vs. 70%; \( p = 0.002 \), particularly in low-risk patients (4-year EFS was 92% in the AATO group \( n = 86 \) vs. 71% in the AIDA group \( n = 92 \); \( p = 0.008 \)). The 4-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 4 years was 27% in the AIDA group and 0% in the AATO group (\( p < 0.0001 \)).

### 3.1.2 Relapsed or Refractory APL

The study by Raffoux et al. [7] compared AATO with ATO, which was not a relevant comparison according to the NICE scope. Therefore, no comparative evidence for ATO in relation to any of the relevant comparators listed in the scope was presented in the CS for this population. OS was similar between the AATO and ATO study arms. Across both arms, the estimated 2-year OS was 59% (95% CI 35–77). EFS was not reported in this study.

### 3.2 Critique of Clinical Effectiveness Evidence and Interpretation

The company conducted systematic reviews of the evidence for ATO 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 guidance.

#### 3.2.1 Newly Diagnosed APL

Of the two trials presented as evidence for untreated APL (APL0406 and AML17), only one (APL0406) was in accordance with the licence. We thus considered this trial to be the main evidence and AML17 as supporting evidence only. A full assessment of the quality of APL0406 by the company and by the ERG was hampered by the fact that only published information was available for assessment as the trial was not conducted by Teva. Overall, the trial appeared to have been well conducted. The effectiveness data showed that relevant patient outcomes were improved. The safety data showed that patients would need to be carefully selected and informed of the particular risks of the chosen regimen. The ERG highlighted that knowledge of long-term toxicity of AATO for newly diagnosed patients awaits a post-authorisation long-term safety cohort study.

The clinical expert from the company stated that “in the UK patients are treated following the AML17 protocol”. However, AML17 also included patients at high risk who did not form part of the population considered in the NICE scope. 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.
3.2.2 Relapsed or Refractory APL

The company presented one trial in relapsed/refractory patients. The trial by Raffoux et al. [7] compared AATO with ATO, which was not a relevant comparison according to the NICE scope. Hence, the ERG considered that the company had not presented any relevant evidence for relapsed or refractory APL. In view of this lack of relevant evidence, the ERG considered that non-RCTs could have been included in the submission for the relapsed/refractory population.

European Medicines Agency 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 this approval [8, 9]. However, these two studies were not included in the company submission.

The ERG also noted that no trials in the CS compared ATO regimens with HSCT or with BSC as specified in the NICE scope.

3.3 Cost-Effectiveness Evidence Submitted by the Company

The company conducted systematic literature reviews to identify relevant cost effectiveness, health-related quality of life and resource use and costs studies. Although the systematic literature reviews identified cost-effectiveness analyses in the literature, the company decided to develop a de novo model. The model structure proposed by the company, however, diverged from those used in the identified cost-effectiveness analyses. The company justified this by stating that the existing cost-effectiveness analyses 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 [1].

The model structure developed by the company considered different treatment phases: first-line, second-line, HSCT (including both allogeneic and autologous HSCT) and other phases (i.e. treatment-related myelodysplastic syndrome or acute myeloid leukaemia and death).

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

As mentioned above, the company only assessed the cost effectiveness of AATO compared with AIDA in the newly diagnosed low- to intermediate-risk APL population, i.e. first-line treatment. The cost effectiveness of ATO in the relapsed/refractory APL population was not assessed in the CS.

AATO was modelled with up to two cycles (of 4 weeks) of induction therapy followed by eight cycles (of 4 weeks) of consolidation therapy. The only comparator considered, first-line AIDA, was implemented with up to two cycles (of 4 weeks) of induction therapy followed by three cycles (of 4 weeks) of consolidation therapy. For both AATO and AIDA, maintenance treatment was not modelled; 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.

The transition probabilities from the first-line states of the model were informed by the APL0406 trial. The sources and justification for the transition probabilities 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 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, utility values 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 and medical costs were mainly obtained from the British National Formulary, NHS reference costs and Personal Social Services Research Unit. NHS reference costs were used to inform the costs of managing adverse events; alternatively, published literature was used. Costs were presented in British pounds (£, 2015 values).

In the CS base-case (probabilistic), AATO was less expensive (£31,088 saved) and more effective (2.546 quality-adjusted life-years (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.

3.4 Critique of Cost-Effectiveness Evidence and Interpretation

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, health-related quality of life, and cost and healthcare resource use studies.

The ERG considered that the model structure was more complex than that used in published cost-effective analyses, and appropriate to reflect this condition and treatment pathway. The main ERG concerns regarding the model structure related to inconsistencies between treatments regarding the modelling of patients that cannot be evaluated for molecular remission, an error in the number of tunnel states used to represent the 2-year molecular remission health state, the absence of disease-related mortality in the ‘on treatment’ health states and the applicability of an allogeneic HSCT health state when modelling the UK clinical practice. These issues were considered in the additional analyses performed by the ERG.

The model time horizon of 40 years resulted 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.

ATO was only assessed in the newly diagnosed APL population (first-line). The company provided an analysis for the relapsed/refractory population (second-line) in its clarification response. However, the company’s description of this analysis did not provide clarity on precisely how this analysis was performed. The ERG therefore implemented their own ‘second-line’ 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 BSC as comparators in the second-line setting. However, the ERG considered 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’s concerns related to the estimation of treatment effectiveness included (1) multiple reference/calculation errors, (2) the overestimation of cardiac events resulting in patients switching to second-line induction for AIDA, (3) assumptions and calculation errors related to the relapse probabilities and (4) 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 informing transition probabilities from the HSCT health states were 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 the 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 that were not extensively discussed in the CS. These issues were considered in the additional analyses performed by the ERG.

The ERG agreed with the company that utility values for APL patients elicited through the EQ-5D were probably not available in the literature. However, the ERG was concerned about the validity of the utility values used in the CS base-case for the following reasons: the selection process of the utility values and the assumptions underlying disutilities associated with adverse events were unclear, their non-adherence with the NICE reference case, and the lack of justification supporting the various 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 Personal Social Services Research Unit and NHS reference costs 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 perceived the expected life expectancy estimated by 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 lifetime horizon, the ERG decided to cap the health state utility values in the model using the general population utility values.
3.5 Additional Work Undertaken by the ERG

In the CS base-case (probabilistic analysis), 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 incorporated various adjustments to the company base-case, which included fixing six errors, adjusting the time horizon, using unadjusted utility values, capping the utility values to not exceed the general population utility values (over time) and alternative assumptions related to the remission probability. This resulted in the (deterministic) ERG base-case, wherein AATO remained dominant. Furthermore, one of the main uncertainties, according to the ERG, related to the (long-term) extrapolation of the relative treatment effect, which is not extensively discussed in the CS. In the CS, it is implicitly assumed that treatment benefits are maintained for the entire time horizon (i.e. constant relative treatment effect). This uncertainty is difficult to parametrise, hence the ERG explored the potential impact of the uncertainty related to the extrapolation by performing an exploratory scenario analysis wherein the relapse probability was assumed to be equal for AATO and AIDA 2 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. In addition, a worst-case scenario was produced to acknowledge the uncertainties highlighted by the ERG. The worst-case scenario was a combination of some of the scenario analyses explored by the ERG, including alternative assumptions regarding the extrapolation of treatment effectiveness and the applicability of allogeneic HSCT to the UK clinical setting. This resulted in an incremental cost-effectiveness ratio (ICER) of £21,622 per QALY gained (deterministic). Due to issues related to the probabilistic analyses in the economic model submitted by the company and the time available, 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 were able to produce probabilistic results for its base-case.

3.6 Conclusion of the ERG Report

Both trials in newly diagnosed APL (APL0406 and AML17) compared AATO with AIDA. 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 18–71 years and took place in Italy and Germany. Results from APL0406 showed that 97.3% of people in the AATO group had not had an event after 50 months, compared with 80.0% in the AIDA group. This difference was statistically significant (p < 0.001). More people having AATO were alive at 50 months compared with people having AIDA (99.2% compared with 92.6%; p = 0.007). There was also a statistically significant (p = 0.001) lower cumulative incidence of relapse with AATO (1.9%) compared with AIDA (13.9%) at 50 months. The primary outcome in AML17 was health-related quality of life. The results did not show a statistically significant difference between AATO and AIDA in most health-related quality-of-life outcomes, but the study may have been underpowered to detect this difference. At 4 years there was a statistically significant difference in EFS (91% with AATO and 70% with AIDA; p = 0.002) but not in OS (93% and 89%; p = 0.250).

The only trial presented for relapsed/refractory patients was by Raffoux et al. [7]. This small trial compared AATO with ATO, which is not a relevant comparison according to the scope.

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 remains dominant in the scenario analyses conducted by the company. The ERG incorporated various adjustments to the company base-case, which resulted in the 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 explore the impact of the identified uncertainties on the results. This resulted in an ICER of £21,622 per QALY gained for AATO vs. AIDA.

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 per QALY gained. However, as indicated by the subgroup analysis performed by the ERG, the ICER of AATO for the second line might be substantially different (estimated to be £31,184 per QALY gained).

4 National Institute for Health and Care Excellence Guidance

On June 13, 2018, NICE recommended ATO, within its marketing authorisation, as an option for inducing remission and consolidation in adults with untreated, low- to intermediate-risk APL (when administered with ATRA), as well as in adults with relapsed or refractory APL (after previous treatment with retinoid and chemotherapy).
4.1 Consideration of Clinical Effectiveness

The AC concluded that study APL0406 was relevant to NHS clinical practice in England because it used the dosing schedule and population defined in the marketing authorisation for ATO. AML17 used a lower dose (about 60% of that in the marketing authorisation) and included people with high-risk disease; therefore, AML17 was considered as supporting evidence.

The AC acknowledged that the results for relapsed or refractory APL would decrease over time. The AC concluded that it could recommend ATO, as an option, for treating relapsed or refractory APL [3].

5 Key Methodological Issues

Uncertainty analyses in decision analytic models, such as probabilistic sensitivity analyses, tend to focus on statistical uncertainty. This is typically limited to what is described, according to the GRADE handbook (for grading the quality of evidence and the strength of recommendations), as imprecision of evidence [10]. However, uncertainty due to indirectness of evidence, for instance due to discrepancies between the population, intervention and/or outcomes of interest and those considered in the evidence, is often not incorporated in probabilistic sensitivity analyses. Indirectness of evidence can arguably have a more pronounced impact on the uncertainty surrounding the (cost) effectiveness than statistical uncertainty. Our critical review of ATO for the treatment of APL illustrates the usefulness of scenario analyses (including worst-case scenarios) to consider the impact of indirectness or even lack of evidence (e.g. extrapolation of the relative effectiveness of ATO, see Sect. 3.5). Neglecting uncertainty caused by indirectness might lead to suboptimal reimbursement decisions. Hence, more emphasis should be placed on the consideration of indirectness in decision analytic models, preferably in the probabilistic sensitivity analysis (PSA). Ideally, distributions should be parameterised to reflect the subjective uncertainty of the decision maker, but usually, sampling distributions based solely on empirical observation are used, which take no account of indirectness. As a second-best solution, and if the uncertainty is difficult to parameterise, scenario analyses can be used for this purpose as, for instance, is done to explore the potential impact of the uncertainty related to the extrapolation of the relative effectiveness of ATO.

Another methodological issue that was raised by the ERG was the inconsistency of LYs and QALYs estimated for the APL population with those of the general UK population (i.e. 30.12 vs. 29.62 [11, 12]). Although this might have face validity in some assessments, in this patient population this is questionable. Hence, age- and sex-matched UK population LYs and QALYs can serve as a general validation benchmark for the LYs and QALYs estimated in decision analytic models. This similarly applies to health-state utility values and the mortality probabilities applied over time.
6 Conclusions

This paper describes the single technology appraisal considering ATO for the treatment of APL. The ERG had multiple methodological reservations regarding the approach to estimate the cost effectiveness. This was acknowledged by the AC, which stated that some model assumptions led to the cost effectiveness analyses being uncertain. However, the AC recognised that, despite being uncertain, for untreated APL the most plausible cost-effectiveness estimate was likely to be < £20,000 per QALY gained and thus ATO was deemed cost effective compared with AIDA. Consistently, for relapsed or refractory APL, the AC stated that ATO is already used and that the clinical- and cost-effectiveness evidence is uncertain because the clinical trial was small and did not compare ATO with AIDA. However, the AC concluded that it is likely that ATO is clinically effective and represents a cost-effective use of NHS resources in relapsed or refractory APL. As a result, ATO is recommended for both untreated and relapsed or refractory APL.

Acknowledgements The authors of this report would like to acknowledge Piet Portegijs for his contribution to the clinical-effectiveness sections of the original ERG report.

Author contributions All authors have commented on the submitted manuscript and have given their approval for the final version to be published. Rob Riemsma, Debra Fayter, Sohan Deshpande, Gill Worthy and Jos Kleijnen reviewed the clinical-effectiveness evidence, Steven Duffy reviewed the search methods. Bram Ramaekers, Sabine Grimm, Willem Witlox, Xavier Pouwels, Nigel Armstrong and Manuela Joore reviewed the cost-effectiveness evidence. Bram Ramaekers acts as overall guarantor for the manuscript. This summary has not been externally peer reviewed by PharmacoEconomics.

Compliance with Ethical Standards

Funding This project was funded by the National Institute for Health Research (NIHR) Health Technology Assessment (HTA) Programme. See the HTA programme website for further project information (http://www.hta.ac.uk). This summary of the ERG report was compiled after NICE issued the Final Appraisal Determination (FAD). Consistently, for relapsed or refractory APL, the AC stated that ATO is already used and that the clinical- and cost-effectiveness evidence is uncertain because the clinical trial was small and did not compare ATO with AIDA. However, the AC concluded that it is likely that ATO is clinically effective and represents a cost-effective use of NHS resources in relapsed or refractory APL. As a result, ATO is recommended for both untreated and relapsed or refractory APL.

Conflict of interest Bram L. T. Ramaekers, Rob Riemsma, Sabine Grimm, Debra Fayter, Sohan Deshpande, Nigel Armstrong, Willem Witlox, Xavier Pouwels, Steven Duffy, Gill Worthy, Jos Kleijnen and Manuela A. Joore have no conflicts directly relevant to the content of this article.

Open Access This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

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. Riemsma R, Ramaekers B, Grimm S, Fayter D, Deshpande S, Portegijs P, et al. Arsenic trioxide for treating acute promyelocytic leukaemia: a Single Technology Assessment [Internet]. York: Kleijnen Systematic Reviews Ltd; 2018. https://www.nice.org.uk/guidance/ta526/documents/committeepapers. Accessed 3 July 2018.
3. National Institute for Health and Care Excellence. Arsenic trioxide for treating acute promyelocytic leukaemia. Final appraisal determination [Internet]. London: National Institute for Health and Care Excellence; 2018. https://www.nice.org.uk/guidance/ta526/documents/final-appraisal-determination-document. Accessed 3 July 2018.
4. 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. https://www.nice.org.uk/guidance/tda10216/documents/finalscope. Accessed 8 Dec 2017.
5. 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.
6. 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 leukaemia: final results of the randomized Italian-German APL0406 trial. J Clin Oncol. 2017;35(6):605–12.
7. 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 leukaemia. J Clin Oncol. 2003;21(12):2326–34.
8. European Medicines Agency. Trisenox (arsenic trioxide): EPAR—scientific discussion [Internet]. London: European Medicines Agency; 2005. https://www.ema.europa.eu/documents/scientific-discussion/trisenox-epar-scientific-discussion_en.pdf. Accessed 26 Sep 2018.
9. European Medicines Agency. Trisenox (arsenic trioxide): EPAR - Product Information. Annex I. Summary of product characteristics [Internet]. London: European Medicines Agency; 2009. https://www.ema.europa.eu/documents/product-information/trisenox-epar-product-information_en.pdf. Accessed 26 Sep 2018.
10. Schünemann H, Brozek J, Guyatt G, Oxman A, editors. GRADE Handbook: handbook for grading the quality of evidence and the strength of recommendations using the GRADE approach. Updated October 2013 [Internet]; 2013. http://gdt.guidelinedevelopment.org/app/handbook/handbook.html. Accessed 23 July 2018.
11. Office for National Statistics (ONS). National Life Tables, England & Wales, 1980-82 to 2014-16 [Internet]. Newport: Office for National Statistics; 2017. https://www.ons.gov.uk/releases/natlifetablesuk2014to2016. Accessed 23 July 2018.
12. 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. https://www.york.ac.uk/media/che/documents/papers/discussionpapers/CH%20Discussion%20Paper%20172.pdf. Accessed 15 Feb 2018.