Azacitidine for Treating Acute Myeloid Leukaemia with More Than 30% Bone Marrow Blasts: an Evidence Review Group Perspective of a NICE Single Technology Appraisal

Running heading **Azacitidine for Treating AML with more than 30% Bone Marrow Blasts**

Irina Tikhonova,[1,3,4] Martin Hoyle,[1] Tristan Snowsill,[1] Chris Cooper,[1] Jo Varley-Campbell,[1] Claudius Rudin,[2] Ruben Mujica Mota[1]

[1] Peninsula Technology Assessment Group (PenTAG), Exeter, UK
[2] Royal Devon and Exeter Hospital, Exeter, UK
[3] I.Tikhonova@exeter.ac.uk
[4] Fax: n/a

Abstract

The National Institute for Health and Care Excellence (NICE) invited the manufacturer of azacitidine (Celgene) to submit evidence for the clinical and cost-effectiveness of this drug for the treatment of acute myeloid leukaemia (AML) with more than 30% bone marrow blasts in adults who are not eligible for hematopoietic stem cell transplantation, as part of the Institute's Single Technology Appraisal (STA) process. The Peninsula Technology Assessment Group (PenTAG) was commissioned to act as the Evidence Review Group (ERG). The ERG produced a critical review of the evidence contained within the company's submission to NICE.

The clinical effectiveness data used in the company's economic analysis was derived from a single randomised controlled trial, AZA-AML-001. It was an international, multicenter, controlled, phase III study with an open-label, parallel-group design, conducted to determine the efficacy and safety of azacitidine against a conventional care regimen (CCR). CCR was a composite comparator of AML treatments currently available in the NHS: intensive chemotherapy followed by best supportive care (BSC) upon disease relapse or progression, non-intensive chemotherapy followed by BSC, and BSC only. In AZA-AML-001, the primary endpoint was overall survival (OS). Azacitidine appeared to be superior to the CCR, with median OS of 10.4 and 6.5 months, respectively. However, in the intention-to-treat analysis, survival advantage associated with azacitidine was not statistically significant.

The company submitted a *de novo* economic evaluation based on a partitioned survival model with four health states: "Remission", "Non-remission", "Relapse/Progressive disease" and "Death". The model time horizon was 10 years. The perspective was NHS and Personal Social Services. Costs and health effects were discounted at the rate of 3.5% per year. The base-case incremental cost-effectiveness ratio (ICER) of azacitidine compared with the CCR was £20,648 per quality-adjusted life-year (QALY) gained. In their probabilistic sensitivity analysis (PSA), the mean ICER was £17,423 per QALY. At the willingness-to-pay of £20,000, £30,000 and £50,000 per QALY, the probability of azacitidine being cost-effective was 0.699, 0.908 and 0.996, respectively.

The ERG identified a number of errors in Celgene's model and concluded that the results of the company's economic evaluation could not be considered robust. After amendments to Celgene's model, the base-case ICER was £273,308 per QALY gained. In the PSA, the mean ICER was £277,123 per QALY. At a willingness-to-pay of £100,000 per QALY, the probability of azacitidine being cost-effective was less than 5%. In all exploratory analyses conducted by the ERG, the ICER exceeded the NICE's cost-effectiveness threshold range of £20,000 to £30,000 per QALY. Given the evidence provided in the submission, azacitidine did not fulfil NICE's End-of-Life criteria.

After considering the analyses conducted by the ERG and submissions from clinician and patient experts, the NICE Appraisal Committee did not recommend azacitidine for this indication.

Keywords: Azacitidine, Acute myeloid leukaemia, AML, Cost-effectiveness, Conventional care regimen, CCR, Single Technology Appraisal.

Key Points for Decision Makers
The cost-effectiveness analysis was based on clinical evidence derived from a single randomised control trial, AZA-AML-001. In the trial, efficacy and safety of azacitidine was compared against a conventional care regimen (CCR) for the treatment of acute myeloid leukaemia (AML) in adults with more than 30% bone marrow blasts, not eligible for hematopoietic stem cell transplantation. The CCR was a composite comparator of AML treatments available within the NHS in England and Wales. In the trial, azacitidine demonstrated a median overall survival benefit vs. the CCR (10.4 vs. 6.5 months). In the intention-to-treat analysis, improvement in overall survival was not statistically significant.
After reviewing the clinical and cost-effectiveness evidence on azacitidine, this treatment appeared to represent poor value for money in England and Wales at a cost-effectiveness threshold of £30,000 per quality-adjusted life-year gained.
The NICE Appraisal Committee did not recommend azacitidine within its licensed indication.

1 Introduction

For a health technology to be recommended for use within the National Health Service (NHS) in England and Wales, its clinical and cost-effectiveness must be demonstrated. The National Institute for Health and Care Excellence (NICE), which is a part of the NHS and an independent organization, is responsible for providing evidence-based guidance on promoting good health and preventing and treating ill health [1].

The NICE Single Technology Appraisal (STA) process is specifically designed to give recommendations, in the form of NICE guidance, on the use of a health technology, within a single indication, in the NHS [2]. The NICE Appraisal Committee makes a judgement on clinical and cost-effectiveness of the technology, based on a pharmaceutical company's submission, a report from an Evidence Review Group (ERG), which is an external academic organization appointed by NICE to appraise the company's submission, and the views of organizations representing healthcare professionals, patients, carers and the government.

Within the STA process, the company provides NICE with a written submission and a mathematical model describing the clinical and cost-effectiveness of the technology in question. The economic evidence, presented in the submission, shows the effectiveness of the treatment in relation to how much it costs the NHS. The ERG's role is to appraise the company's submission in consultation with clinical specialists, and prepare a report with additional analyses exploring alternative assumptions relevant to clinical and cost-effectiveness of the technology.

After consideration of all the relevant evidence, the Appraisal Committee formulates preliminary guidance, an Appraisal Consultation Document (ACD), recommending or not recommending the intervention. The stakeholders are invited to comment on the ACD and the submitted evidence. A subsequent ACD may be produced or a Final Appraisal Determination (FAD) issued. The submission of a Patient Access Scheme (PAS) by the company sponsoring the technology is possible in order to improve the cost-effectiveness of the medicine and allow NICE to recommend treatment that would otherwise not have been found to be cost-effective.

NICE invited the manufacturer of azacitidine (Celgene) to submit evidence for the clinical and cost-effectiveness of this drug for the treatment of acute myeloid leukaemia (AML) with more than 30% bone marrow blasts in adults who are not eligible for hematopoietic stem cell transplantation (HSCT), as part of the Institute's Single Technology Appraisal (STA) process. The Peninsula Technology Assessment Group (PenTAG) was commissioned to act as the Evidence Review Group (ERG). The ERG prepared a report to NICE, which gives a thorough critique of the submission by Celgene. This manuscript presents a summary of the company's submission, a summary of the report prepared by the ERG, and the NICE guidance. All the relevant appraisal documents are available on the NICE website [3].

2 The Decision Problem

AML is a cancer of the bone marrow [4]. It affects myeloid blood cells which perform a number of important functions such as fighting bacterial infections, defending the body against parasites and preventing the spread of tissue damage. In AML, immature blood-forming cells accumulate in the bone marrow, blood, and organs, and interfere with the production of normal blood cells. AML can arise *de novo*, due to the progression of other diseases (e.g., myelodysplastic syndrome, MDS) or due to exposure to cytotoxic agents [5]. More than half (55%) of new cases of AML are diagnosed in people aged 70 and over [6]. In England, the annual incidence rate of AML in people of 65 years and older is about 17/100,000 [7]. Although AML is a relatively rare disease, its incidence is expected to increase as the population ages.

Age and cytogenetics appear to be the most important prognostic factors. The 5-year survival for AML patients less than 65 years old is 41.6% and just 5.4% in patients of 65 years and older [8]. Cytogenetic status in AML patients is generally classified as being favourable, intermediate or poor, with survival differing markedly between these groups. In a recent study, median overall survival for patients aged over 60 was 1.6 years for individuals in the favourable group, 0.9 years for individuals in the intermediate group, and 0.5 years for

individuals in the group with poor cytogenetics [9]. It has also been shown that AML secondary to MDS is associated with a worse prognosis [10].

AML progresses rapidly and may lead to fatal infections, bleeding or organ infiltration within months if left untreated [5]. Treatment with curative intent requires an aggressive chemotherapy regimen. The first phase of the treatment, an induction therapy, targets leukemic blasts, with the aim of attaining a complete remission. Once complete remission is achieved, a consolidation therapy is usually given to prevent relapse and eradicate any residual or remaining leukemic cells from the blood and bone marrow. HSCT offers the best means of preventing AML recurrence [11]. However, this procedure is associated with high treatment-related morbidity and mortality in older patients, and therefore not recommended.

In older patients, it is more difficult to achieve a remission, and relapse is more common. One reason is the high risk of toxicity during intensive chemotherapy treatment. Besides, patients in this age group often have other medical problems, such as central nervous system conditions and impaired kidney or liver function, and, as a result, may not be able to withstand the required doses of treatment. Therefore, intensive chemotherapy is restricted to patients with a favourable performance status, minimal organ dysfunction and without severe comorbid diseases. For other patients, treatment options currently available in the NHS consist of non-intensive chemotherapy or best supportive care (BSC) only. A number of guidelines provide recommendations on the treatment of elderly patients with AML [12, 13]. However, there is no formal risk algorithm used in the UK clinical practice to distinguish patients eligible for intensive vs. non-intensive approaches; treatment choices take into account the features of the disease, presence of comorbidities, performance status as well as patient's preference.

In 2011, NICE approved a cytostatic anticancer drug, azacitidine, for use in the NHS as a first-line treatment for adults who are not eligible for HSCT and have:
- intermediate-2 and high-risk MDS according to International Prognostic Scoring System
- chronic myelomonocytic leukaemia with 10-29% marrow blasts without myeloproliferative disorder
- AML with a blast count of 20-30% and multi-lineage dysplasia according to the World Health Organisation's classification (NICE TA218) [14]

On 28th October, 2015, the European Medicines Agency granted marketing authorization to extend the licensed indication for azacitidine to include AML patients aged ≥65 years with >30% marrow blasts, not eligible for HSCT. The objective of this technology appraisal was to evaluate the clinical and cost-effectiveness of azacitidine according to its recent licensed indication. Comparators selected for this appraisal were: intensive chemotherapy (IC) with anthracycline in combination with cytarabine; non-intensive chemotherapy with low-dose cytarabine ara-c (LDAC); and BSC only including blood product replacement, antibiotics, antifungals and intermittent low dose chemotherapy with hydroxycarbamide. The following measures of clinical effectiveness were considered relevant for this appraisal: overall survival (OS), progression-free survival (PFS), time to disease progression, response rates (including haematologic response and improvement), blood-transfusion independence, infections, treatment-emergent adverse events (TEAEs) and health-related quality of life (HRQoL). The scope developed by NICE for this assessment specified that the economic analysis should follow the NICE reference case [15].

3 The ERG's Review

3.1 Clinical Evidence Provided by the Company

The clinical-effectiveness evidence in the company's submission was based on data from one randomised control trial (RCT), AZA-AML-001 (NCT01074047) [16]. It was an international, multicenter, controlled, phase III study with an open-label, parallel-group design, which evaluated efficacy and safety of azacitidine vs. a conventional care regimen (CCR) in 488 patients age ≥65 years with *de novo* or secondary AML with >30% bone marrow blasts and an Eastern Cooperative Oncology Group Performance Status (ECOG PS) of 0 to 2 with adequate organ function, who are not eligible for HSCT. The CCR was a composite comparator consisting of IC followed by BSC upon disease relapse or progression, LDAC followed by BSC, and BSC only.

Before randomization, an individual CCR (IC+BSC, LDAC+BSC or BSC only) was preselected for each patient on the basis of age, ECOG PS, comorbidities, and regional guidelines and/or institutional practice. Once the CCR had been chosen, a central, stratified and permuted block randomization method, and interactive voice response system were used to randomly assign patients to receive azacitidine with BSC upon disease relapse or progression (N=241) or the preselected CCR (N=247; IC+BSC=44, LDAC+BSC=158, BSC only=45). Following randomization and the therapy treatment, follow-up appointments were scheduled once per week during the first two treatment cycles, and every other week thereafter. The frequency of safety and efficacy measures ranged from weekly to every 12 weeks, depending on the procedure. Drug administration and data collection protocols are outlined in Table 1 (Online Resource). The patients had a follow-up visit for the collection of adverse events up to 28 days after the last dose of trial drug or up to the end-of-study visit, whichever period was longer. After this visit, the patients were followed for survival on a monthly basis until death, lost to follow-up, withdrawal of consent or end of the study. The duration of the study was 31 months: a 19-month study enrolment period, followed by 12 months of treatment and observation.

In AZA-AML-001, the primary efficacy endpoint was overall survival. Secondary outcomes are shown in Table 2 (Online Resource). The median duration of follow-up was 24.5 months. During the trial, no crossover between any treatment groups was allowed. However, patients who stopped the study treatment could receive subsequent AML therapy during study follow-up. A total of 69 patients (28.6%) in the azacitidine group and 75 patients (30.4%) in the CCR group received subsequent therapy. The most common therapies received in the azacitidine and the CCR groups included a cytarabine-based regimen (16.6% and 11.3%, respectively), azacitidine (4.6% and 13.0%, respectively), and/or decitabine (0.8% for each group) [16]. By the study end, there were 193 deaths (80.7%) following treatment with azacitidine and 201 deaths (81.4%) following the CCR treatment.

Azacitidine was superior to the CCR in prolonging patients' survival, with median OS of 10.4 (95% confidence interval [CI]: 8.0-12.7 months) and 6.5 months (95% CI: 5.0-8.6 months), respectively. However, the intention-to-treat analysis showed that azacitidine was not statistically superior to the CCR (stratified hazard ratio [HR] = 0.85; 95% CI: 0.69-1.03). Refer to Table 3 (Online Resource) for further details. Celgene justified the failure of the primary endpoint by an imbalance in patients' baseline characteristics and prognostic factors, and the impact of subsequent therapies, which resulted in an underestimation of the true treatment effect of azacitidine vs. the CCR.

Fig. 1 (a) OS for the intention-to-treat population
Median OS was 10.4 months (95% CI: 8.0-12.7 months) for the azacitidine arm and 6.5 months (95% CI: 5.0-8.6 months) for the CCR arm. In the analysis stratified by ECOG PS and cytogenetic risk, the HR was 0.85 (95% CI: 0.69-1.03; log-rank P = .1009). One-year survival was 46.5% for the azacitidine arm and 34.2% for the CCR arm (difference, 12.3%; 95% CI: 3.5%-21.0%). Median follow-up for OS was 24.4 months. There were 193 deaths in the azacitidine arm (80.1%) and 201 deaths in the CCR arm (81.4%).

(b) OS censored for subsequent AML therapy
A total of 67 azacitidine patients and 75 CCR patients were censored at the time they received subsequent AML therapy. Median OS was 12.1 months (95% CI: 9.2-14.2 months) for the azacitidine arm and 6.9 months (95% CI: 5.1-9.6) for the CCR arm. In the analysis stratified by ECOG PS and cytogenetic risk, the HR was 0.76 (95% CI: 0.60-0.96; log-rank P = .0190). CIs for the difference in 1-year survival probabilities were derived by using Greenwood's variance estimate.

AML acute myeloid leukaemia, *CCR* conventional care regimen, *CI* confidence interval, *ECOG PS* Eastern Cooperative Oncology Group Performance Status, *HR* hazard ratio, *OS* overall survival
(+) Censored patient
Source: individual patient data from AZA-AML-001 trial, reproduced with permission from the submission.

In a pre-specified analysis, censoring patients who received AML treatment after discontinuing study drug, median OS with azacitidine vs. the CCR was 12.1 months (95% CI: 9.2-14.2 months) vs. 6.9 months (95% CI: 5.1-9.6 months) [16]. The Kaplan-Meier (KM) plot of time to death from any cause is shown in Fig. 1, and a summary of OS is presented in Table 3 (Online Resource). KM curves for event-free survival (EFS), relapse-

free survival (RFS) among those with a partial or complete response, and progression-free survival (PFS), for other patients, are shown in Fig. 1 (Online Resource). Median EFS was 6.7 months for the azacitidine arm and 4.8 for the CCR arm (HR = 0.87; 95% CI: 0.72-1.05; P = 0.1495); median RFS was 9.3 months and 10.5, respectively (HR = 1.11; 95% CI: 0.75-1.66; P = 0.5832). PFS was not reported in the trial and was estimated from other end-points (for further details, refer to Fig. 1, Online Resource). Measures of haematologic response, duration of remission and RFS were similar between treatment arms when the CCR was combined. However, when comparing azacitidine with the individual CCRs, IC appeared superior to azacitidine for these outcomes, although the study was not powered to detect such differences [3].

Subgroup analyses for patients with poor cytogenetics and patients with MDS-related changes were included in the company's submission. Median OS for people with a baseline cytogenetic risk rated as poor was 6. 4 months in the azacitidine group compared with 3. 2 months in the CCR group (HR = 0.68; 95% CI: 0.68-0.94, P = 0.0185); median OS for patients with MDS-related changes was 12. 7 months and 6. 3 months, respectively (HR 0.69; 95% CI 0.48 to 0.98, P = 0.357).

According to the summary of product characteristics, azacitidine is most commonly associated with haematological reactions including thrombocytopenia, neutropenia and leukopenia (usually grade 3 to 4) and gastrointestinal events including nausea, vomiting (usually grade 1 to 2) or injection site reactions. The company reported that in the trial azacitidine was generally well tolerated. When adjusted for time of exposure, the incidence rates of adverse events in the azacitidine group were lower when compared with the combined care regimen treatments. Grade 3 or 4 TEAEs which occurred in ≥10% of study participants in any treatment group are presented in Table 4 (Online Resource).

The company used the European Organisation for Research Treatment of Cancer (EORTC) QLQ-C30 questionnaire to assess HRQoL [17]. The questionnaire was completed at baseline, on day one of every other cycle and at the end-of-study visit. Primary HRQoL endpoints reported include fatigue score, dyspnoea, physical functioning and global health status. Both regimens were associated with general improvements in health-related quality of life in these QLQ-C30 domains. Statistical analysis of the HRQoL data was not presented in the submission. For further details on HRQoL, refer to Appendix (Online Resource).

3.1.1 Adjustments of Overall Survival Estimates for Subsequent Therapy

In order to address the problem of confounding effects of subsequent active treatments on overall survival, the company performed survival adjustment using various statistical methods. One of them was the inverse-probability-of-censoring-weights (IPCW) method [18] whereby data for switchers were censored at the point of switch while remaining observations were weighted by the inverse probability of not being censored. The company performed two different analyses using IPCW method: one with adjustment for subsequent treatment applied to both trial arms, azacitidine and the CCR, and the other with adjustment applied to the CCR arm only. The results of the latter analysis were adopted for the company's base case. Other methods, such as the Rank Preserving Structural Failure Time Model (RPSFTM) [19] and the Iterative Parameter Estimation (IPE) [20], were also employed. However, they were considered inferior to the IPCW.

Since the difference in overall survival between the arms in AZA-AML-001 trial narrowed over time (Fig. 1, a), and statistically insignificant survival improvement offered by azacitidine was reversed when data were censored at initiation of subsequent AML treatment (Fig. 1, b), Celgene undertook further analyses to adjust for imbalance in the use of subsequent treatment, and baseline covariates across the trial arms. Various methods were explored including Cox proportional hazards (PH) and IPCW Cox PH methods [3]. Methods of adjusting for subsequent treatment produced similar estimates of hazard ratio of azacitidine vs. the CCR. For example, in the baseline covariate-unadjusted IPCW Cox PH method, HR was 0.77 (95% CI: 0.61–0.98), and in the Cox PH unadjusted for baseline characteristics and censoring at switch to AML therapy, HR was 0.75 (95% CI: 0.59–0.95). In contrast, adjusting for baseline covariates had a large effect on the results, e.g., HR in the adjusted Cox PH censoring at switch was 0.69 (95% CI: 0.54–0.88; P = 0.0027).

HR estimates reported by the company relied on the assumption of constant proportional hazards, which was statistically tested for the Cox but not for the IPCW approaches. The company provided no reasons for the absence of tests of the proportional hazards assumption in the IPCW analyses.

3.1.2 Critique of the Clinical Evidence and Interpretation

A clinical effectiveness systematic review, carried out by the company, did not miss any relevant evidence. AZA-AML-001 trial was appropriately designed, although it was underpowered for comparisons of azacitidine with each of the individual comparators.

In the trial, a significant proportion of patients underwent subsequent active treatments, including those not currently offered in the NHS. This was a potential limitation of the study design for the purposes of informing NICE decision, since subsequent therapies were not balanced between treatment arms, which resulted in confounded estimates of the primary effectiveness outcome and other endpoints.

The clinical-effectiveness analysis, conducted by Celgene, adjusted for subsequent treatment with azacitidine in the CCR arm. However, limited results of adjusting for subsequent treatments in the azacitidine arm were presented by the company, based on an incorrect interpretation of technical guidance on adjustment for treatment switching [21]. This misinterpretation led the company to base their main analyses on effectiveness estimates not adjusted for subsequent treatment in the azacitidine arm, which was an important limitation.

The ERG requested the individual patient data to perform the following analyses: a) to replicate the OS analysis conducted by the company, including the IPCW adjustment for treatment switching in both arms with and without controlling for baseline covariates; b) to test the proportional hazard assumption in the IPCW adjusted OS analysis that formed the basis of the company's economic base case; and c) to identify the best parametric survival function for the cost-effectiveness model. Unfortunately, data provided by the company allowed the ERG to replicate only the company's base-case analyses. Thus, statistical analyses of time-to-event outcomes relied on the proportional hazards assumption, which was not justified in the company's submission.

The distribution of patients over the individual CCR therapies, which was observed in the trial and used in the company's base-case analysis, appeared different to current practice in the NHS. The company acknowledged that, and carried out a scenario analysis, using relevant data from a registry in Yorkshire.
A detailed critique of clinical evidence submitted by the company can be found in the ERG's report [22].

Fig. 2 Model structure

3.2 Cost-Effectiveness Evidence Provided by the Company

The company conducted a systematic literature review of the cost-effectiveness of different treatment strategies for the target population of AML patients, which did not find any pre-existing studies addressing the decision problem. The company, therefore, submitted a *de novo* economic evaluation based on a partitioned survival model with four health states: "Remission", "Non-remission", "Relapse/Progressive disease" and "Death" (Fig. 2).

The model starts after all patients have completed the induction cycle of either azacitidine or the CCR treatments. Those patients, who have responded to the treatment by the end of the induction cycle and have complete remission (CR) or complete remission with incomplete blood count recovery (CRi), enter the model in the "Remission" state; patients with partial remission or stable disease are placed in the "Non-remission" state. In the model, patients remain in the "Remission" and "Non-remission" states until disease relapse or progression, upon which they move to the "Relapse/Progressive disease" state. Transition to "Death" may occur from any state.

The model has two treatment arms: azacitidine and the CCR (composed of IC + BSC, LDAC + BSC and BSC only). In both arms, patients are assumed to receive the relevant first-line treatment followed by BSC (upon disease relapse or progression), or BSC only. The model cycle length of 4 weeks corresponds to one treatment

cycle of azacitidine. The model time horizon is 10 years, which, according to the company, covers the remaining lifetime of most patients in the study population. The model perspective was NHS and Personal Social Services. Costs and quality-adjusted life years (QALYs) were discounted at 3.5% per annum.

The proportion of patients in each health state was estimated using RFS, PFS and OS data from the AZA-AML-001 trial. Various parametric models were fitted to the data, and the model selection process proposed by the NICE Decision Support Unit [23] was followed, which resulted in the selection of an exponential model for OS, a Weibull model for RFS and a Gompertz model for PFS. The IPCW estimate of OS effect, which was selected for the company's base-case cost-effectiveness analysis, was obtained assuming proportional hazards. Hazard ratio estimates were derived from analysis adjusting for subsequent treatment with azacitidine in patients randomised to the CCR. OS in azacitidine arm as well as EFS, RFS and PFS in both arms were not adjusted for subsequent treatment.

Health state utility values were estimated by mapping EORTC QLQ-C30 data, collected in the AZA-AML-001, to EQ-5D utility values, and were independent of treatment. The impact of adverse events on HRQoL was accounted for by applying disutility weights to adverse events of grade 3 or 4 from the AZA-AML-001 [25].

The model incorporated costs associated with AML treatment, management of TEAEs of severity grade 3 or above, transfusion costs, tests to monitor disease, and care at the end of life. Treatment costs included drug acquisition, administration and dispensing for azacitidine and other drugs. Drug acquisition costs were estimated using the average daily dose in AZA-AML-001 trial and list prices (British National Formulary [24]), with a confidential PAS discount on the cost of azacitidine. In the base case, full wastage (i.e., no vial sharing across days or across patients) was assumed. Resource use for drug administration, medical management, diagnostic tests and transfusions were estimated from a questionnaire conducted by Celgene with 7 clinicians, who were asked to provide estimates of these quantities according to disease stage (remission, non-remission, and relapse/progressive disease). The estimates of unit costs and the costs of TEAEs treatment were based on the Personal Social Services Research Unit (PSSRU) costs of health and social care [25], and the NHS reference costs 2013 to 2014 [26].

In the company's base-case analysis, the incremental cost-effectiveness ratio (ICER) for azacitidine vs. the CCR was £20,648 per QALY gained. The largest cost components in the azacitidine and the CCR arms were drug acquisition and drug administration costs, respectively. In addition, two subgroups were modelled: patients with poor cytogenetics, and patients with MDS-related changes. For these subgroups, analyses were performed without adjustment for subsequent active treatment. The respective ICERs were £20,227 and £19,175 per QALY gained.

Sensitivity analyses were performed to explore uncertainty in the ICER, and identify parameters to which the model predictions were sensitive. Deterministic sensitivity analysis showed that the ICER was most sensitive to administration costs in the CCR arm, the hazard ratio for OS, the remission rates in the CCR arm, and the acquisition and administration costs in the azacitidine arm. In the probabilistic sensitivity analysis (PSA), the mean ICER was £17,423 per QALY. At willingness-to-pay of £20,000, £30,000 and £50,000 per QALY, azacitidine was cost-effective, vs. the CCR, in 69.9%, 90.8% and 99.6% of iterations, respectively.

Scenario analyses were also conducted to explore the impact of structural uncertainty on the estimates of cost-effectiveness (see the company's submission for further details [3]).

3.2.1 Critique of the Cost-Effectiveness Evidence and Interpretation

The ERG considered the economic evaluation, submitted by Celgene, from the perspective of the NICE reference case [15]. This section discusses the key issues identified by the ERG; further details are provided in the report [22].

The model assumed that no patients would receive active treatment following discontinuation of the study drug. However, in the AZA-AML-001 trial, about 30% of study participants received active second-line treatment (see section 3.1). Therefore, the model assumption on the absence of subsequent treatment was inconsistent with

the data. Clinical advice received by the ERG, confirmed that active second-line treatment is considered for some AML patients in the NHS.

OS in the azacitidine arm was not adjusted for subsequent active treatment. This resulted in an inconsistency between the modelled costs and health outcomes, since only the costs of BSC following discontinuation of azacitidine treatment were modelled.

The model assumed proportional hazards for OS. However, this assumption has no support in the trial data. The ERG believes that the treatment effects employed in Celgene's model may be biased because of the untested and implausible assumption of proportional hazards in the IPCW OS analysis that informed the company's base case.

The ERG noted significant differences in healthcare costs associated with the "Relapse/Progressive disease" state in the azacitidine and the CCR arms, though patients in both arms are expected to receive BSC at this point.

Numerous issues were identified in the implementation of the model in Microsoft Excel. For details, see the ERG report [22]. In summary, they related to the formula used to calculate healthcare resource use and the extrapolation of outcomes. The most significant of those was an error in the modelling of the duration of first-line treatment, which resulted in underestimation of the drug acquisition and administration costs in both arms.

3.3   Additional Work Undertaken by the ERG

3.3.1   Survival Analyses

Individual-patient data, provided by Celgene on request from the ERG, allowed the latter to perform censor-at-switch analysis for subsequent AML therapy in both trial arms, as well as testing for the proportional hazards assumption in the survival data. A range of parametric curves including proportional hazard (exponential, Weibull, Gompertz, bathtub) and non-proportional hazard (log-normal and log-logistic) parametric models were fitted to OS data from each trial arm. For unadjusted analyses, according to the goodness-of-fit (Akaike Information Criterion and Bayes' Information Criterion) and survival model parameter test statistics, only Gompertz model (consistent with PH assumption), fitted to OS data from the CCR arm, was supported by the data, while the exponential was the best model fit to OS data in the azacitidine arm. As for models that adjusted for baseline covariates, the exponential was the optimal model for both arms. The HR for the exponential model was 0.65 and had a predicted difference in OS of 3.64 months, in favour of azacitidine. The baseline covariates included in adjustment were those fixed covariates used in Celgene's IPCW analysis: age, sex, ECOG, cytogenetic risk, preselected CCR treatment, comorbidity group, AML days, platelet transfusion status, geographical region [22].

The ERG's preferred base-case analysis adopted the exponential OS HR estimates adjusted for baseline covariates. Since finding the optimal fitting functional form for RFS and PFS is highly uncertain due to very small number of cases of progression and relapse after initiation of subsequent treatment, the ERG adopted the Kaplan-Meier curves from AZA-AML-001 for these outcomes.

Notably, the adjusted censored-at-switch analysis, performed by the ERG to obtain estimates of OS effectiveness, was not extended to subgroup analysis to avoid overfitting due to small sample size. The latter was based on the original (unadjusted) estimates of censor-at-switch exponential OS used by the company in their model. Further details are provided in the ERG report [22].

3.3.2   Cost-effectiveness Analyses

The ERG identified and corrected model errors. The cumulative effect of those corrections was to increase the company base-case ICER from £20,648 to £62,518 per QALY gained. After all amendments to the model, the ERG's preferred base-case ICER was £273,308 per QALY. The major drivers of the increase in the ICER were:

- the correction to assume the same treatment costs in the "Relapse/Progressive disease" state across treatment arms
- the calibration of the mean number of treatment cycles in the model to match those from the AZA-AML-001 trial
- correction of model errors

A detailed account of these changes is provided in Tables 5-6 (Online Resource).

Fig. 3 Tornado diagram of the ERG's preferred base-case deterministic sensitivity analysis

AZA azacitidine, *AE* adverse event, *CCR* conventional care regimens, *CR* complete remission, *CRi* complete remission with incomplete blood count recovery, *ERG* Evidence Review Group, *HR* hazard ratio, *ICER* incremental cost-effectiveness ratio, *OS* overall survival, *PD* progressive disease, *PFS* progression-free survival, *PR* partial response, *QALY* quality-adjusted life-year, *RFS* relapse-free survival, *SD* stable disease, *TEAEs* treatment emergent adverse events

Source: the ERG's report [25]

The ERG performed sensitivity analyses for their preferred base case. Plausible variations of parameter values in the deterministic sensitivity analysis resulted in ICERs of above £200,000 per QALY (Fig. 3). According to the results, the ICER is most sensitive to the administration costs of azacitidine.

Fig. 4 Cost-effectiveness acceptability curves from the ERG's preferred base-case probabilistic sensitivity analysis

CCR conventional care regimens, *ERG* Evidence Review Group, *QALY* quality-adjusted life-year

The mean ICER of £277,123 per QALY was obtained in a PSA conducted by the ERG. Parameters were varied according to their sampling distributions: beta distribution for response rate, health state utility values, utility values for AEs, and incidence of AEs; and gamma distribution for patients' weight and height, drug usage, number of treatment cycles, and healthcare resource use. At a willingness-to-pay of £100,000/QALY the probability of azacitidine being cost-effective was less than 5% (Fig. 4).

A number of additional analyses were carried out.

Two scenarios explored the effect on the ICER of uncertainty in health resource use estimates obtained from the survey conducted by Celgene: an extreme scenario assuming no healthcare costs in the "Relapse/Progressive disease" state in both treatment arms; and another scenario with no inpatient hospitalization costs, incurred in this disease state, which represented the main difference in costs between the azacitidine and the CCR arms.

The ERG also explored the effect of adopting Celgene's assumption on the costs of monitoring tests and transfusions. In Celgene's analyses, these costs were measured till the end of treatment, while ERG's base-case analysis included these costs for the whole duration of the "Remission" and "Non-remission" phases of the model as well as for the "Relapse/Progressive disease" phase.

The ERG also conducted exploratory subgroup analyses by preselected CCR treatment and explored the effect on ICER of applying a patient distribution over individual CCRs, observed in the NHS.

In all additional analyses conducted by the ERG, the ICER exceeded £30,000 per QALY gained.

3.3.3 End-of-Life Criteria

The company derived their estimated extension of life expectancy of 3.8 months, offered by azacitidine, from median OS estimates for azacitidine and CCR patients from the pivotal trial. The ERG considered the estimate neither plausible nor robust. According to the ERG's analyses, based on the restricted mean survival at 30 months, treatment with azacitidine offers an extension to life of 1.8–2.5 months (further details are provided in Table 7, Online Resource). The results suggest that azacitidine does not fulfil NICE's End-of-Life criteria.

3.4 Conclusions of the ERG Report

On the basis of the clinical evidence provided in the submission, azacitidine appeared to be more clinically effective than the combined comparator comprised of AML treatments currently available in the NHS. The ERG could not obtain a robust estimate of cost-effectiveness of azacitidine using the submitted version of the company's model, and had to make several corrections. After amendments to the model, azacitidine was not cost-effective at a threshold of £30,000 per QALY gained.

4   Key Methodological Issues

Several methodological issues were highlighted during the appraisal. The most important of them was the approach used by Celgene for adjusting survival estimates for subsequent AML treatment.

In Celgene's base-case analysis, OS adjustment for subsequent treatment was performed in the CCR arm only. However, in the AZA-AML-001 trial patients from both arms received second-line treatment (see section 3.1). The company justified their approach by citing methodological guidelines from NICE DSU Technical Support Document 16 [21].

The ERG believes that the company misinterpreted the guidelines, and, in their base-case analysis, overall survival estimates in both arms should have been adjusted for subsequent active treatment.

5   NICE Guidance

In March 2016, the NICE Appraisal Committee reviewed the data available on the clinical and cost-effectiveness of azacitidine, having considered evidence on the nature of AML and the value placed on the benefits of azacitidine by people with the condition, those who represent them and clinical experts. The Committee took into consideration the effective use of NHS resources. Following further consultation, NICE issued its guidance on the use of azacitidine for AML: azacitidine was not recommended for AML with > 30% bone marrow blasts in people of 65 years or older not eligible for HSCT. NICE published the Committee's decisions for this appraisal. Stakeholders were given the opportunity to appeal against the Committee's recommendations. NICE received no appeals, and the Final Guidance was published on 27 July, 2016.

5.1 Consideration of Clinical and Cost-Effectiveness Issues

This section discusses the key issues considered by the Appraisal Committee. The full list can be found in the FAD [3].

The Committee expressed concerns about the company's decision problem in which the individual regimens were combined into the single conventional care regimen, CCR. It was noted that in work done for previous NICE appraisals, the NICE decision support unit advised against the use of such 'blended' comparator. The Committee concluded that the relevant comparators for the appraisal were those specified in the NICE scope.

The Committee discussed the approaches used by the company to adjust for treatment switching. It noted that they were all susceptible to bias because they relied on the proportional hazards assumption, which the Committee did not consider appropriate. The Committee accepted the changes made by the ERG to use Kaplan–Meier curves from each trial arm rather than fitting curves that assumed proportional hazards.

The Committee discussed the approaches employed by the company and the ERG for modelling costs in the "Relapsed/Progressive disease" state. The methodology used by the company meant that in this health state, different resource use estimates were applied to the azacitidine and the CCR groups despite the fact that patients from both groups were treated with best supportive care. The Committee concluded that the assumption of the same cost of "Relapse/Progressive disease" state in both arms, employed by the ERG, was justified. However, the Committee noted that the ERG had equalised the cost to the higher estimate of resource use, and that taking the average of the resource use estimates across all groups might have been more reasonable.

The Committee considered the modelling of the number of treatment cycles. It heard from the ERG that treatment duration in Celgene's model was inconsistent with AZA-AML-001 trial. The Committee agreed with the ERG that the number of treatment cycles in the model should reflect AZA-AML-001.

The Committee discussed the implementation errors in the company model identified by the ERG. The company agreed with all error corrections made by the ERG. The Appraisal Committee accepted all the amendments proposed by the ERG and concluded that the base-case ICER in the company's model was £63,000 per QALY gained.

The Committee accepted the ERG's restricted mean overall survival estimate of 2.5 months and concluded that azacitidine was unlikely to meet the criteria to be considered a life-extending End-of-Life treatment.

6   Conclusions

On the basis of the evidence submitted by the company, the ERG concluded that azacitidine was not cost-effective at the NICE's threshold range of £20,000 to £30,000 per QALY for treating AML with more than 30% bone marrow blasts in people of 65 years or older who are not eligible for hematopoietic stem cell transplantation. After considering the analyses conducted by the ERG and submissions from clinician and patient experts, the NICE Appraisal Committee did not recommend azacitidine for this indication.

Contributions of Authors
I.T. drafted the manuscript; M.H., T.S., C.C., J.V.-C. C.R. and R.M.M. revised the manuscript for intellectual content prior to submission. R.M.M. is the overall guarantor of the content.

Conflict of interest
I.T., M.H., T.S., C.C., J.V.-C. C.R. and R.M.M. declared no conflict of interest.

CHEERS checklist
Table 8 (Online Resource) contains CHEERS checklist [28] which gives references to where in the manuscript or Online Resource the CHEERS checklist items are satisfied.

1. National Institute For Health and Care Excellence. NICE Charter. 2013. https://www.nice.org.uk/Media/Default/About/Who-we-are/NICE_Charter.pdf. Accessed 13 July 2016.
2. National Institute for Health and Care Excellence. Guide to the single technology appraisal process. 2009. https://www.nice.org.uk/Media/Default/About/what-we-do/NICE-guidance/NICE-technology-appraisals/Guide-to-the-single-technology-appraisal-process.pdf. Accessed 6 Apr 2016 2016.

3. National Institute for Health and Care Excellence. Azacitidine for treating acute myeloid leukaemia with more than 30% bone marrow blasts. NICE technology appraisal guidance [TA399]. 2016. https://www.nice.org.uk/guidance/ta399. Accessed 29 July 2016.

4. National Health Service. Acute myeloid leukaemia. 2016. http://www.nhs.uk/conditions/Leukaemia-acute/Pages/Introduction.aspx. Accessed 6 Apr 2016.

5. Kumar CC. Genetic abnormalities and challenges in the treatment of acute myeloid leukemia. Genes Cancer. 2011. doi:10.1177/1947601911408076.

6. Cancer Research UK. Acute myeloid leukaemia (AML) incidence statistics. 2011-2013. http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/leukaemia-aml/incidence. Accessed 6 Apr 2016.

7. Epidemiology and Cancer Statistics Group at the University of York DoHS, The University of York. Haematological Malignancy Research Network. Incidence. https://www.hmrn.org/statistics/incidence Accessed 6 Apr 2016.

8. National Cancer Institute Surveillance, Epidemiology, and End Results Program. Actute myeloid leukemia, 5-year relative survival (percent) 2003-2009 by age at diagnosis. http://seer.cancer.gov/archive/csr/1975_2010/browse_csr.php?sectionSEL=13&pageSEL=sect_13_table.16.html#table3. Accessed 6 Apr 2016.

9. Mrozek K, Marcucci G, Nicolet D, Maharry KS, Becker H, Whitman SP et al. Prognostic significance of the European LeukemiaNet standardized system for reporting cytogenetic and molecular alterations in adults with acute myeloid leukemia. J Clin Oncol. 2012. doi:10.1200/JCO.2012.43.4738.

10. Oran B, Weisdorf DJ. Survival for older patients with acute myeloid leukemia: a population-based study. Haematologica. 2012. doi:10.3324/haematol.2012.066100.

11. Appelbaum F. The use of bone marrow and peripheral blood stem cell transplantation in the treatment of cancer. CA-Cancer J Clin. 1996;46(3):142-64.

12. Doehner H, Estey EH, Amadori S, Appelbaum FR, Buechner T, Burnett AK et al. Diagnosis and management of acute myeloid leukemia in adults: recommendations from an international expert panel, on behalf of the European LeukemiaNet. Blood. 2010. doi:10.1182/blood-2009-07-235358.

13. 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.

14. NICE. Azacitidine for the treatment of myelodysplastic syndromes, chronic myelomonocytic leukaemia and acute myeloid leukaemia. 2011. https://www.nice.org.uk/guidance/ta218. Accessed 6 Apr 2016.

15. National Institute for Health and Care Excellence. Guide to the methods of technology appraisal. 2013. http://publications.nice.org.uk/pmg9. Accessed 6 Apr 2016.

16. Dombret H, Seymour JF, Butrym A, Wierzbowska A, Selleslag D, Jang JH et al. International phase 3 study of azacitidine vs conventional care regimens in older patients with newly diagnosed AML with >30% blasts. Blood. 2015. doi:10.1182/blood-2015-01-621664.

17. Aaronson NK, Ahmedzai S, Bergman B, Bullinger M, Cull A, Duez NJ et al. The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. Journal of the National Cancer Institute. 1993;85(5):365-76.

18. Robins JM, Finkelstein DM. Correcting for noncompliance and dependent censoring in an AIDS Clinical Trial with inverse probability of censoring weighted (IPCW) log-rank tests. Biometrics. 2000;56(3):779-88.

19. Robins JM, Tsiatis AA. Correcting for non-compliance in randomized trials using rank preserving structural failure time models. . Commun Stat Theory. 1991;20(8):2609-31.

20. Branson M, Whitehead J. Estimating a treatment effect in survival studies in which patients switch treatment. Statistics in Medicine. 2002;21(17):2449-63.

21. Latimer NR, Abrams KR. NICE DSU Technical Support Document 16: Adjusting survival time estimates in the presence of treatment switching. 2014. http://www.nicedsu.org.uk. Accessed 13 July 2016.

22. Mujica Mota R, Varley-Campbell J, Tikhonova I, Cooper C, Hoyle M, Rudin C et al. Azacitidine for treating acute myeloid leukaemia with more than 30% bone marrow blasts: A Single Technology Appraisal. 2016.

23. NICE. NICE DSU Technical Support Document 14: Survival analysis for economic evaluations alongside clinical trials - extrapolation with patient-level data. 2013. http://www.nicedsu.org.uk/NICE%20DSU%20TSD%20Survival%20analysis.updated%20March%202013.v2.pdf. Accessed 7 Apr 2016.

24. Joint Formulary Committee. British National Formulary 69. The authority on the selection and use of medicines. 2015:588.

25. Curtis L. Unit costs of health and social care. Canterbury: Personal Social Services Research Unit (PSSRU); 2014.

26. Department of Health. NHS reference costs 2013 to 2014. London: DH2014.

27. National Institute for Health Research. Health Technology Assessment (HTA) Programme. 2016. http://www.hta.ac.uk. Accessed 8 Apr 2016.

28. Husereau D, Drummond M, Petrou S. et al. PharmacoEconomics. 2013. 31: 361. doi:10.1007/s40273-013-0032-y.