Azacitidine for treating acute myeloid leukaemia with more than 30% bone marrow blasts

A Single Technology Appraisal

Produced by	Peninsula Technology Assessment Group (PenTAG) University of Exeter St Luke's Campus, Heavitree Road, Exeter, EX1 2LU
Authors	Ruben Mujica Mota,[1] *Senior Lecturer* Jo Varley-Campbell,[1] *Associate Research Fellow* Irina Tikhonova,[1] *Associate Research Fellow* Chris Cooper,[1] *Senior Information Specialist* Martin Hoyle,[1] *Associate Professor* Claudius Rudin,[2] *Consultant Haematologist* Tristan Snowsill,[1] *Research Fellow* [1] Peninsula Technology Assessment Group (PenTAG), Exeter, UK [2] Royal Devon and Exeter NHS Foundation Trust, Exeter, UK
Correspondence to	Ruben Mujica Mota, South Cloisters, St Luke's Campus, Heavitree Road, Exeter, EX1 2LU
Date completed	10/02/2016
Source of funding	This report was commissioned by the NIHR HTA Programme as project number 15/64/10.
Declared competing interests of the authors	None
Acknowledgments	We acknowledge the excellent administrative support of Sue Whiffin and Jenny Lowe (both of University of Exeter). We also acknowledge support and guidance given by Mary Bond (University of Exeter) relating to the critique of the company's clinical evidence.
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:	Mujica Mota R, Varley-Campbell J, Tikhonova I, Cooper C, Hoyle M, Rudin C, Snowsill T. Azacitidine for treating acute myeloid leukaemia with more than 30% bone marrow blasts: A Single Technology Appraisal. Peninsula Technology Assessment Group (PenTAG), 2016.

Contributions of authors

Ruben Mujica Mota	Led the critique of the review of economic evaluation studies, the critique of the economic model and subsequent treatment adjustment. Wrote the respective sections of these contributions.
Jo Varley-Campbell	Led the critique of the company's decision problem and clinical effectiveness evidence. Wrote the Decision problem and Clinical effectiveness chapters. Contributed to the writing and editing of the report.
Irina Tikhonova	Contributed to the critique of the company's economic model and to the writing and editing of the Cost-effectiveness chapter.
Chris Cooper	Led the critique of the company's literature searching for this submission. Wrote the review of the literature searches for the report. Contributed to the writing and editing of the report.
Martin Hoyle	Provided occasional advice on the critique of the economic evaluation, commented on drafts of the report and is the guarantor of the report.
Claudius Rudin	Provided clinical advice on acute myeloid leukaemia and its management within the NHS. Reviewed and revised a draft version of the report.
Tristan Snowsill	Contributed to the critique of the company's clinical and cost-effectiveness evidence. Wrote the Summary, Background, End of Life and Overall conclusions sections. Compiled the report. Provided overall project management.

Contents

List of tables

List of figures

Abbreviations

AE	Adverse event
AIC	Akaike information criterion
AML	Acute myeloid leukaemia
ANC	Absolute neutrophil count
AZA	Azacitidine
BIC	Bayesian information criterion
BM	Bone marrow
BNF	British National Formulary
BSC	Best supportive care
CCR	Conventional care regimens
CI	Confidence interval
CMML	Chronic myelomonocytic leukaemia
CNS	Clinical nurse specialist
CR	Complete remission
CRc	Cytogenetic complete remission
CRi	Complete remission with incomplete blood count recovery
DNA	Deoxyribonucleic acid
ECOG	Eastern Cooperative Oncology Group
EFS	Event-free survival
EORTC	European Organization for Research and Treatment on Cancer
EQ-5D	EuroQol five dimensions questionnaire
ERG	Evidence Review Group
HMRN	Haematological Malignancy Research Network
HR	Hazard ratio

HRQoL	Health-related quality of life
HCRU	Healthcare resource utilisation
HSCT	Haematopoietic stem cell transplantation
IC	Intensive chemotherapy
ICER	Incremental cost-effectiveness ratio
IPCW	Inverse probability of censoring weighting
IPE	Iterative parameter estimation
ITT	Intent-to-treat
IV	Intravenous
IVRS	Interactive Voice Response System
IWG	International Working Group
KM	Kaplan Meier
LDAC	Low-dose cytarabine
LYG	Life years gained
MDS	Myelodysplastic syndromes
NICE	National Institute for Health and Care Excellence
OS	Overall survival
PAS	Patient Access Scheme
PD	Progressive disease
PFS	Progression-free survival
PH	Proportional hazards
PR	Partial remission
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PRO	Patient-reported outcomes
PS	Performance status
PSA	Probabilistic sensitivity analysis
PSSRU	Personal Social Services Research Unit

QALY	Quality-adjusted life year
QLQ	Quality of life questionnaire
RBC	Red blood cell
RCT	Randomised controlled trial
RFS	Relapse-free survival
RPSFTM	Rank preserving structural failure time model
SAE	Serious adverse event
SC	Subcutaneous(ly)
SD	Stable disease
SF-12	Short-form 12 questionnaire
WBC	White blood cell
WHO	World Health Organization

1 Summary

1.1 Critique of the decision problem in the company submission

The company narrowed the population from adults with acute myeloid leukaemia (AML) and bone marrow blasts more than 30% (as per the NICE Scope) to adults aged ≥65 years who are not eligible for haematopoietic stem cell transplantation with AML and bone marrow blasts more than 30% to adults. This change was to coincide with the European Medicines Agency marketing authorisation for azacitidine and was deemed a reasonable change by the ERG.

The intervention in the decision problem was azacitidine, as in the NICE Scope.

The comparator(s) in the decision problem were different from the NICE Scope. The company replaced three individual comparators (intensive chemotherapy [IC], non-intensive chemotherapy with low dose cytarabine [LDAC] and best supportive care [BSC]) with one composite comparator (conventional care regimen; CCR) on the basis that there are no established criteria for selecting one CCR. As a result, the company has not assessed whether azacitidine demonstrated clinical and/or cost-effectiveness versus each of the CCR comparators. The ERG considered this to be a weakness of the submission.

The company reported the same outcomes to that of the NICE Scope.

The NICE Scope asked for evidence, if available, on the following subgroups: people with AML secondary to myelodysplastic syndrome (MDS) and people with adverse-risk cytogenetics. The company reported that these subgroups were assessed. Although the submission looked at the subgroup of AML with MDS-related changes (which is a broader subgroup than AML secondary to MDS), these other considerations were deemed acceptable by the ERG.

1.2 Summary of clinical effectiveness evidence submitted by the company

The primary focus of the company's submission was the RCT AZA-AML-001. Patients were randomised to azacitidine (N=241) or to a conventional care regimen (N=247; BSC=45, LDAC=158, IC=44). Baseline characteristics were reported as being balanced between arms. Outcome results were as follows:

Overall survival

Azacitidine was numerically superior to CCR in prolonging survival of adults ≥65 years with AML with >30% bone marrow blasts but statistical significance was not reached. Median duration of follow up was 24.5 months. By the study end, there were 193 deaths (80.7%) following treatment with azacitidine and 201 deaths (81.4%) following CCR treatment.

Secondary endpoints

1-year survival rates were 46.5% for azacitidine compared to 34.3% in the CCR arm (difference 12.3 %; 95% CI: 3.5, 21.0).

Measures of haematologic response, duration of remission and remission free survival were similar between treatment arms when CCR was combined. When CCR was not combined, it

appeared that IC was numerically superior to azacitidine for these outcomes, although the study was not powered to detect any such differences.

No statistical analyses were presented for the health-related quality of life (HRQoL) data. Appearances from the figures suggest that CCR was favourable to azacitidine.

Adverse events

Treatment related AEs were common for both azacitidine, LDAC and IC. Unsurprisingly, AEs were less common for BSC.

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

The company presented a poorly constructed systematic review of the literature. Their searches were weak and their inclusion criteria were both over- and under-exclusive. Ultimately however, the ERG concluded that the company did not miss any evidence.

The primary focus of the company's submission was the RCT AZA-AML-001. This was generally an appropriately-designed RCT, although it was underpowered for comparisons of azacitidine to each of the CCR arms. It is not clear whether the proportion of patients pre-selected to each CCR therapy in the RCT (18% IC, 64% LDAC and 18% BSC) are representative of NHS clinical practice; data from a registry in Yorkshire suggests more patients may receive BSC [redacted] and fewer LDAC [redacted], while clinical expert advice is that more patients would be expected to receive IC. The use of subsequent therapies following treatment assignment was permitted, and this was a limitation to the study design as it resulted in confounded estimates for the primary efficacy endpoint and other endpoints.

The open-label design of the trial, although unavoidable as the treatments generally require different levels of medical intervention, increases the risk of bias.

Statistical analyses of time-to-event outcomes relied on the proportional hazards assumption, which transpired not to be justified.

1.4 Summary of cost-effectiveness evidence submitted by the company

1.4.1 Company's systematic review of economic evaluations

The company conducted a systematic review of economic evaluations, which did not find any pre-existing studies adequately addressing the decision problem.

1.4.2 Company's submitted economic evaluation

1.4.2.1 Methods

The company presented a model-based economic evaluation to address the decision problem.

A semi-Markov (survival partition) model was used with four health states: Remission, Stable disease, Relapse/Post-progression and Death. Patients achieving remission started the model in the Remission state, while patients not achieving remission started in the Stable disease state. A model cycle length of four weeks was used, and a time horizon of ten years

was used. Outputs of the model (costs, life years and quality-adjusted life years [QALYs]) were discounted at 3.5% per annum.

Two treatment arms were modelled. The azacitidine (AZA) arm modelled treatment with azacitidine until discontinued, followed by BSC. The CCR arm modelled a mixture of conventional treatments (IC, LDAC and BSC), with IC and LDAC followed by BSC after discontinuation.

Overall survival, relapse-free survival and progression-free survival curves were constructed by fitting parametric survival models to data from the AZA-AML-001 trial. The treatment effect was modelled using proportional hazards for all survival curves. A model selection process was followed, which resulted in the selection of an exponential survival model for overall survival, a Weibull model for relapse-free survival and a Gompertz model for progression-free survival. Hazard ratios of 0.84 and 0.85 were used for relapse-free and progression-free survival respectively, while a hazard ratio of [redacted] was used for overall survival based on an analysis adjusting for subsequent treatment with azacitidine in patients randomised to CCR.

Health state utility values were estimated by mapping EORTC QLQ-30 data collected in the AZA-AML-001 trial to EQ-5D utility values, and were not modelled as varying according to treatment given. The impact of adverse events on health-related quality of life was also directly modelled by treatment.

Costs were modelled from the NHS and personal social services perspective. Drug acquisition costs were estimated using the average daily dose in AZA-AML-001 and list prices (British National Formulary; BNF), with a confidential patient access scheme (PAS) discount of [redacted] applied to the cost of azacitidine. In the base case full wastage was assumed (i.e., no vial sharing across days or across patients). Patients were assumed to receive the relevant first-line treatment until relapse or progression. Drug administration, medical management, diagnostic test and transfusion resource use were estimated through a survey of clinicians conducted by the company. The PSSRU Unit cost of health and social care and the NHS reference costs were used to estimate unit costs. Costs of adverse events were also modelled.

Univariate and probabilistic sensitivity analyses were conducted to explore uncertainty in the incremental cost-effectiveness ratio (ICER) and to identify parameters to which the model was sensitive. Scenario analyses were also conducted.

1.4.2.2 Results

In the company's base case analysis, treatment with CCR resulted in 0.6365 QALYs and £40,608 cost, while treatment with azacitidine resulted in [redacted] QALYs [redacted] and [redacted] cost [redacted], with a corresponding ICER of £20,648 per QALY.

Azacitidine was predicted to provide QALY gains across all health states, and was predicted to result in increased costs in the Remission and in Relapse/Progressive disease health states, partially compensated for by savings in the Stable disease health state.

Drug acquisition costs were the largest cost component in the AZA arm [redacted] more costly than in the CCR arm), while drug administration costs were the largest cost component in the CCR arm [redacted] more costly than in the AZA arm). Other costs were largely similar between the two arms.

In their probabilistic sensitivity analysis, incremental QALYs were similar to the deterministic analysis, while incremental costs were marginally lower. The resulting ICER for azacitidine was £17,423 per QALY. At cost-effectiveness thresholds of £20,000, £30,000 and £50,000 per QALY, azacitidine was cost-effective versus CCR in 69.9%, 90.8% and 99.6% of iterations respectively.

Univariate sensitivity analyses identified that the results were sensitive to a number of parameters, with administration costs in the CCR arm, the hazard ratio for overall survival, the remission rates in the CCR arm and the acquisition and administration costs in the AZA arm as the five parameters to which the model was most sensitive.

One notable scenario analysis showed that when overall survival in the CCR arm was not adjusted for subsequent treatment, this resulted in improved cost-effectiveness of azacitidine (£11,537 per QALY).

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

The ERG identified several issues with the company's submitted economic evaluation.

The model assumed that no patients would receive active treatment following discontinuation of first-line treatment. However, in the AZA-AML-001 trial underpinning the analysis, 29% of participants received active second-line treatment. Advice from clinical experts suggests that active second-line treatment is considered for some patients in the NHS.

The model assumed proportional hazards for all time-to-event outcomes, even though this was not supported for overall survival and relapse-free survival by results from the AZA-AML-001 trial.

Overall survival in the AZA arm was not adjusted for subsequent active treatment, resulting in an inconsistency between the modelled health outcomes and costs, since only the costs of best supportive care were modelled following azacitidine.

Implementation issues were identified in the model. The most significant of these was an error in the calculation of the duration of first-line treatment which resulted in an underestimate of the drug acquisition and administration costs in both arms.

The ERG also identified that there were significant differences in the cost associated with the Relapse/progressive disease state between the AZA and CCR arm, even though all patients (in both arms) are expected to be receiving BSC at this point.

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

1.6.1 Strengths

The company's submission was based on a recent and relevant RCT (AZA-AML-001) with the following strengths:

- Multicentre RCT conducted across multiple countries, including countries in Western Europe;

- Sufficient follow-up for mature estimates of survival outcomes, including overall survival;
- Appropriate dosing in the intervention and comparator arms;
- Appropriate randomisation, including concealment of allocation prior to randomisation and stratified blocking on key prognostic variables;
- Appropriate and relevant outcomes measured.

The company's submitted economic evaluation had the following strengths:

- Simple and transparent overall model structure;
- Inclusion of relevant costs from an NHS and personal social services perspective;
- Suitable cycle length and time horizon;
- Transparent process for fitting survival models;
- Utility values suitably mapped from health-related quality of life measurements from AZA-AML-001;
- Probabilistic sensitivity analysis to estimate the importance of parameter uncertainty in the decision problem.

1.6.2 Weaknesses and areas of uncertainty

The decision problem addressed by the company's submission had a key weakness that instead of individual conventional care regimens as comparators (as in the NICE Scope), a combined comparator was used.

The company's systematic review of clinical effectiveness evidence was hampered by poorly designed and reported searches.

The pivotal trial (AZA-AML-001) had the following weaknesses:

- Underpowered for comparisons between azacitidine and individual conventional care regimens;
- Significant proportion of patients used subsequent active treatments, including treatments not currently used in the NHS, and these were not balanced between treatment arms;
- Limited proportion of patients allocated intensive chemotherapy as their conventional care regimen compared to expectation of routine clinical practice according to clinical experts;
- Open-label design increases the risk of bias;
- Statistical analyses relying on proportional hazards assumption which is not justified.

The company's submitted economic evaluation had the following weaknesses:

- Inconsistency between the treatments costed post-discontinuation (BSC only) and the subsequent treatments reflected in overall survival estimates (AZA: active treatments; CCR, active treatments except azacitidine);

- Inappropriate proportional hazards assumption for overall survival and relapse-free survival outcomes;
- Implementation errors, including a significant underestimate of treatment duration in both arms;
- Inadequate exploration of structural uncertainty;
- Significantly greater costs after relapse/progression in the CCR arm despite patients in both arms receiving BSC only in this state.

The following areas of uncertainty remain:

- The overall survival benefit demonstrated in AZA-AML-001 did not reach statistical significance in pre-planned analyses, yet it is interpreted nevertheless as a positive result;
- It is not clear to what extent azacitidine is a clinically effective and cost-effective alternative to IC, LDAC and BSC as individual comparators.

1.7 Summary of exploratory and sensitivity analyses undertaken by the ERG

The ERG preferred base case ICER is £169,606 per QALY, compared to the company's base case ICER of £20,648 per QALY. The reasons for the increased ICER are:

- Corrections to errors in model formulae (increases ICER from £20,648 to £62,518 per QALY);
- Increased costs in the AZA and CCR arms due to correcting the implementation of treatment duration in the model (increases ICER from £62,518 to £131,698 per QALY);
- Equalised costs in the Relapse/progressive disease health state across the model (increases ICER from £131,698 to £238,674 per QALY);
- Overall survival in both arms adjusted for subsequent active treatment (reduces ICER from £238,674 to £171,511 per QALY);
- Relapse-free survival Kaplan-Meier curves used for AZA and CCR arms (increases ICER from £171,511 to £174,205 per QALY);
- Progression-free survival Kaplan-Meier curves used for AZA and CCR arms (increases ICER from £174,205 to £246,488 per QALY);
- Adjusting overall survival for baseline covariates (reduces ICER from £246,488 to £169,606 per QALY).

2 Background

2.1 Critique of company's description of underlying health problem

Celgene describe acute myeloid leukaemia (AML) as *'an aggressive, clonal myeloid neoplasm with maturation arrest of myelopoiesis, leading to an accumulation of myeloblasts in the [bone marrow (BM)] and/or blood.'* (Source: Celgene submission, Section 3.1, p. 31).

AML is a haematological cancer affecting the myeloid line of blood cells. In AML, myeloid stem cells in the bone marrow produce immature blood cells (usually myeloblasts) which do not develop fully and build up in the bone marrow. These immature blood cells are not able to function properly and they reduce the ability of the bone marrow to produce other cells the body needs.

AML can develop following myelodysplastic syndrome (MDS), or can develop as a result of therapy (e.g., cytotoxic therapy), or can arise without previous associated disease or treatment (primary AML).

The World Health Organisation (WHO) system requires involvement of at least 20% of blood and/or bone marrow by myeloblasts for AML diagnosis, and is also used to classify AML into subtypes to aid clinical decision making and prognosis.

Celgene describe the following signs and symptoms of AML (Source: Celgene submission, Section 3.1, p. 31):

> *The clinical signs and symptoms of AML are diverse and non-specific, but they are usually directly attributable to the leukaemic infiltration of the [bone marrow], with resultant cytopenias (reduction in blood cell counts). Typically, patients present with signs and symptoms of fatigue, haemorrhage, and/or infections and fever due to reductions in [red blood cells], platelets, and [white blood cells].[1] The corresponding impact on physical and psychological aspects of quality of life is significant and increases over the course of the condition.[2]*

The ERG believes the description given is appropriate.

2.1.1 Epidemiology

Celgene give the following estimates of the incidence of AML (Source: Celgene submission, Section 3.1, p. 31):

> *AML is the most frequent form of leukaemia, accounting for approximately 25% of all leukaemia cases in adults in the Western world.[3] […] In the UK between 2009 and 2011, an average of 40% of cases were diagnosed in men and women aged 75 years and over, and almost three quarters of cases (73%) were diagnosed in those aged 60 and over.[4] The median age of diagnosis is between 65 and 72 years for the entire population, and 78 years when evaluating the population who are aged over 65 years.[5-10]*
>
> *The annual incidence rate of AML in England has been estimated to be 4.1 per 100,000.[11] The incidence increases dramatically with older age, rising to 18.35 per*

> *100,000 in people aged 65 years and over,[11] equating to approximately 1,777 new cases of AML in this patient group in England annually.[11-13]*

The incidence statistics provided by Celgene appear to be well-sourced, but we estimate a marginally lower number of new cases in adults aged over 65 per year in England (1,610 versus 1,777) using the same datasets but a different method (applying the different age-specific rates to the relevant population estimates and aggregating afterwards rather than applying an aggregated rate to an aggregate population estimate). This also corresponds to a marginally lower incidence rate for the over-65s of 16.88 per 100,000 (rather than 18.35).

2.1.2 Diagnosis

Diagnostic criteria are given by Celgene (Source: Celgene submission, Section 3.1, p. 31):

> *Diagnosis of AML requires the examination of peripheral blood and BM specimens, using morphology, cytochemistry, immunophenotyping, cytogenetics, and molecular genetics. According to the WHO classification of myeloid neoplasms, a myeloid neoplasm with ≥20% blasts in the peripheral blood or BM is considered to be AML when occurring de novo, evolution to AML when it occurs with previous diagnosis of MDS or myelodysplastic/myeloproliferative neoplasm.[14]*

An abnormal result on a complete blood count is a typical finding prior to a diagnosis of AML. An excess of white blood cells is commonly seen, and counts for platelets and/or red blood cells may be reduced.

2.1.3 Prognostic markers and risk factors

Celgene have provided general information on survival in AML (Source: Celgene submission, Section 3.4, p. 34):

> *AML is a heterogeneous disease in terms of response to treatment and OS. Prognostic factors that contribute to this heterogeneity can be patient-related (such as increased age, reduced performance status, comorbidities, vulnerability, or frailty) or disease-related (such as genetic factors, adverse cytogenetics, somatic mutations, or whether the patient has MDS-related changes).[15-17]*
>
> *Survival is highly age dependent with survival rates being significantly lower in older patients.[5] The median [overall survival (OS)] of elderly patients with AML in population-based studies has remained unchanged since 1995 at 1.5 to 3 months.[18, 19] Furthermore, a recent analysis of the [Haematological Malignancy Research Network] HMRN registry highlights the current poor outcomes in UK routine practice, with a median OS of [redacted] months for non-transplant-eligible AML patients 65 years or older treated with [conventional care regimens (CCR)].[20] There is also a clear disparity in 5-year survival rates between AML patients of different ages. Between 2003 and 2009, 5-year survival rates for patients <65 years of age was 41.6%, but just 5.4% in patients ≥65 years of age.[21] In contrast, the life expectancy of people in the general population once they have reached 75 years of age is a further 10.6 years (males) and 12.9 years (females).[22] Therefore, AML represents a challenging disease to treat, and results in a significant reduction in patient's life expectancy.*

Age and cytogenetics appear to be the most important prognostic factors. The NICE Scope identified two subgroups to be considered if evidence allowed, these were: people with AML secondary to myelodysplastic syndrome, and people with adverse-risk cytogenetics.

AML secondary to myelodysplastic syndrome is associated with reduced likelihood of treatment response and therefore with worse prognosis.[9]

Cytogenetics are generally classified as being favourable, intermediate or poor, with survival differing markedly between these groups. In a recent analysis of patients enrolled in the Cancer and Leukaemia Group B first-line trials, median (overall) survival for patients aged over 60 was 1.6 years for individuals in the favourable group (based on cytogenetics and molecular genetics), 0.9 years for individuals in the intermediate groups, and 0.5 years for individuals in the adverse groups.[23]

2.1.4 Burden and impact on quality of life

Celgene note the following in relation to the impact on the quality of life of the patient (Source: Celgene submission, Section 3.2, p. 32):

> *When compared with the general population, patients with AML experience a significant reduction in physical functioning (as determined via the physical component domain of quality of life assessments), and experience a higher incidence of depression.[2] Furthermore, quality of life deteriorates over time, with a significant reduction observed as early as 2 weeks after AML diagnosis.[2] Patients with AML can also experience appetite loss and fatigue; both having a negative impact on overall measures of quality of life.[24] The burden of the disease continues until death, with patients frequently suffering from open bleeding, infection, and pain during the final stages of the disease.[25]*

The ERG note that the study by Sekeres et al.[2] found similar SF-12 mental component scores for AML patients as population norms, which should be considered alongside the finding of a higher incidence of depression. Other than this the description is appropriate and relevant.

Celgene also note the potential impact on caregivers (Source: Celgene submission, Section 3.2, p. 32):

> *The impact is far reaching with caregivers, including family or friends, often having to deal with numerous and concurrent stressful events, and often suffering negative psychological, behavioural and physiological effects on their daily lives and their health.[26]*

The ERG note that the publication by Bevans and Sternberg[26] cited by the company is a case study of a single individual with AML secondary to MDS, who also received HSCT. The ERG considers that this does not constitute high-quality evidence of an impact on caregivers, although the ERG does not dispute that such impacts may exist.

2.2 Critique of company's overview of current service provision

The company gives the following overview of the clinical pathway of care for *elderly* patients with AML – the company's decision problem focuses on elderly patients as opposed to all adults, *see Section 3 (p. 23)* – (Source: Celgene submission, Section 3.3, p. 33):

> *Due to the heterogeneity of disease, there is no standard of care for elderly patients with AML, resulting in complex treatment guidelines.[27-30] Despite differences between published treatment guidelines, there is a general consensus that treatment decisions should be based on a number of patient- and disease-related prognostic factors. Patients with favourable prognostic factors are more likely to be assessed as "fit" to receive treatment with IC while patients with unfavourable prognostic factors, such as increased age, poor performance and/or cytogenetic risk status, and increased comorbidities are typically deemed unfit for treatment with IC. As such, these patients are usually offered less intensive chemotherapy options, such as LDAC and those unable to tolerate chemotherapy or who chose not to receive LDAC should receive BSC only.*
>
> *Despite this general guidance there is no widely accepted risk algorithm which clinicians use in the UK when deciding which patients are most likely to benefit from intensive or non-intensive treatment options. A recent review further demonstrated the lack of structure when making treatment decisions, concluding that decisions remain complex and selection is subjective based on the clinician's judgement.[31] Patient choice was also found to be a confounding factor, accounting for approximately 8% of treatment decisions, irrespective of the clinicians' recommendation.[32]*

The company also provide the following information in relation to current service provision (Source: Celgene submission, Section 3.7, p. 37):

> *Treatment options for elderly patients with >30% BM blasts AML include HSCT, IC, low-dose chemotherapy (LDAC), or BSC alone.[28] However, HSCT is rarely used in patients older than 65 years.[33] Decitabine is also licenced in the EU for the treatment of elderly WHO-defined AML but it is not reimbursed (NICE TA270[34]) and so is not used in UK routine clinical practice. Treatment with IC is typically contraindicated for patients aged ≥65 years with an adverse performance status, organ damage, and comorbidities.[28] Treatment with IC can however be successfully used in older patients, if restricted to patients with a favourable performance status, minimal organ dysfunction and/or comorbidity, and favourable cytogenetics, but is associated with an increased risk of treatment-related mortality.[5, 28] In this patient population, treatment options usually consist of LDAC or BSC and patients suffer from low survival rates, with a 26% 30-day mortality reported in patients receiving low-intensity treatment.[21, 28]*

The ERG note that a Swedish registry study covering 98% of Swedish patients diagnosed with AML by the French-American-British criteria (requiring at least 30% bone marrow blasts) found that 55% of patients aged at least 65 years with ECOG PS 0–2 were reported

fit for intensive chemotherapy, and 45% of patients aged at least 65 years of any ECOG PS.[22]

Our clinical expert (CR) advised that IC is administered in an inpatient setting, while LDAC may be self-administered at home. Azacitidine would most likely be administered as a regular day case in the NHS, which presents the problem of administering azacitidine on seven consecutive days since many day case units do not currently operate over the weekend. Patients treated with azacitidine would normally attend for five consecutive days (Monday to Friday) and then two consecutive days after a weekend break (Monday and Tuesday). Prevailing clinical opinion is that this schedule is non-inferior to administration over seven consecutive days. Patients treated with best supportive care may be admitted at times to treat certain complications (e.g., infection). A UK study recently found that of patients with one of three haematological malignancies (AML, diffuse large B-cell lymphoma and myeloma), 74% died in hospital, 15% at home, and the remainder in a hospice or nursing home.[35]

The ERG believes that these descriptions are appropriate and relevant to the company's chosen decision problem (*see Section 3, p. 23*).

3 Critique of company's definition of the decision problem

The company presented their decision problem within the Executive Summary chapter, under the subheading 'statement of the decision problem' (Celgene submission, Section 1.1, p. 12–16). A summary table of the NICE Scope,[36] the company's decision problem and the ERG's critique is presented below (*Table 1*). Further comments to the decision problem follow the table.

Table 1: Summary table of decision problem critique

Decision problem	NICE Scope	Company's decision problem	ERG notes
Population	Adults with acute myeloid leukaemia with bone marrow blasts more than 30%	Adults aged ≥65 years who are not eligible for HSCT with AML with >30% marrow blasts	The changes in population by the company brought the population in line with the EMA marketing authorisation for azacitidine. The ERG is satisfied this is a reasonable change.
Intervention	Azacitidine	As per Scope	No comments.
Comparator	• Intensive chemotherapy with an anthracycline in combination with cytarabine • Non-intensive chemotherapy with low dose cytarabine • Best supportive care which may include blood product replacement, antibiotics, antifungals and intermittent low dose chemotherapy with hydroxycarbamide	• Conventional care regimen (CCR; consisting of IC, LDAC and BSC)	The company have replaced three individual comparators with one composite comparator on the basis that there are no established criteria for selecting one CCR. As a result the company have not assessed whether azacitidine demonstrates clinical and cost effectiveness compared to each CCR (in patients for whom that CCR would be appropriate). The ERG considers this to be a weakness of the submission.
Outcome	The outcome measures to be considered include: • Overall survival • PFS • Time to disease progression • Response rates, including haematologic response and improvement • Blood-transfusion independence • Infections • Adverse effects of treatment • Health-related quality of life • Cost per quality-adjusted life year.	The outcomes measured include: • Overall Survival • PFS – estimated from EFS and RFS for the purpose of economic modelling • Time to disease progression • Response rates, including CR, CRc, and PR • Blood-transfusion independence • Infections • Adverse effects of treatment • Health-related quality of life • Cost per quality-adjusted life year.	Two outcomes (PFS and response rate) reported by the company do not match exactly to the Scope. Differences are either terminology or added detail for clarification. These differences are deemed acceptable.
Other considerations	If the evidence allows the following subgroups will be considered. These include: • People with AML secondary to myelodysplastic syndrome • People with adverse-risk cytogenetics	A number of pre-defined patient- and disease-related subgroups were assessed during the pivotal trial, AZA-AML-001 and included those with MDS-related changes, and poor cytogenetic risk status, as per Scope	AML secondary to MDS is a subgroup of AML with MDS-related changes (constituting just over half), but outcomes are expected to be similar.

Key: AML, acute myeloid leukaemia; BM, bone marrow; BSC, best supportive care; CCR, conventional care regimens; CR, complete remissions; CRc, cytogenetic complete remission; EFS, event-free survival; EMA; European Medicines Agency; HSCT; haematopoietic stem cell transplantation; IC, intensive chemotherapy; IV, intravenous; LDAC, low-dose cytarabine; MDS, myelodysplastic syndromes; OS, overall survival; PFS, progression-free survival; PR, partial remission; RFS, relapse-free survival; SC, subcutaneous.

Source: NICE Scope [36] and Celgene submission, Table 1, p. 12–16

3.1 Population

The population in the company's submission did not match for age and eligibility (adults aged ≥65 years who are not eligible for haematopoietic stem cell transplantation), to the population specified in the NICE Scope (adults).[36] Celgene justify this inconsistency by stating that:

> *This submission specifically evaluates the efficacy and tolerability of azacitidine in patients aged ≥65 years and who are not eligible for HSCT in line with the new indication approved by the EMA.*

The European Medicines Agency (EMA) indication for azacitidine has been expanded, and now includes: "Treatment of adult patients aged 65 years or older who are not eligible for HSCT with AML with >30% marrow blasts according to the WHO classification." Therefore, the population reported by the company matches the EMA marketing authorisation for azacitidine but does not match the NICE Scope. Overall we agree that the population considered by the company's submission is appropriate based on it matching the EMA indication.

3.1.1 Subgroups to be considered

The NICE Scope states that if evidence allows, the following subgroups should be considered:

- People with AML secondary to myelodysplastic syndrome;
- People with adverse-risk cytogenetics.

The company's decision problem suggests that pre-defined subgroups were assessed in the pivotal trial, including those with MDS-related changes and those with poor cytogenetic risk status. The ERG note that AML with MDS-related changes is a broader category including AML secondary to MDS, and that in the pivotal trial there were 158 patients with MDS-related changes but only 87 patients with prior MDS.[37] Nevertheless, the prognosis of AML with MDS-related changes is likely to be similar to the prognosis of AML secondary to MDS.

3.2 Intervention

The company's decision problem specified the intervention as 'azacitidine', which matches the NICE Scope.[36]

The NICE Scope describes azacitidine as follows; 'Azacitidine (Vidaza, Celgene) is an analogue of nucleotide cytidine that reduces DNA methylation by inhibition of DNA methyltransferase. Azacitidine is administered subcutaneously.'

The EMA recommend the following for administering azacitidine:

> *The recommended starting dose of Vidaza is 75 mg per square metre body surface area (calculated using the patient's height and weight). It is given as an injection under the skin… …every day for one week, followed by three weeks with no treatment. This four-week period is one 'cycle'. Treatment continues for at least six cycles and then for as long as it benefits the patient. The liver, kidneys and blood should be checked before each cycle. If the blood counts fall too low or if the patient*

> *develops kidney problems, the next treatment cycle should be delayed or a lower dose should be used. Patients who have severe liver problems should be carefully monitored for side effects, but Vidaza must not be used in patients with advanced liver cancer.*

Our clinical advisor (CR) commented that they would typically administer azacitidine for five days (Monday to Friday), and then administer the remaining two days on Monday and Tuesday of the following week, as this is more convenient for the patients and the day case setup at the hospital.

3.3 Comparators

The comparators in the submission do not match those in the Scope.[36] There is more detail given by the Scope, regarding component treatments, than by the company who just report the overarching term for the comparator. The company also refer to decitabine, a treatment that was included in the NICE Draft Scope as a form of non-intensive chemotherapy. Decitabine has marketing authorisation in the UK but has not received a positive NICE recommendation following termination of NICE technology appraisal 270.[34] The company note that decitabine was removed for the Final Scope and do not include it in their decision problem.

The comparator in the company submission is a composite comparator, containing the comparators as identified in the Scope. This comparator is referred to as Conventional Care Regimen (CCR). Further details on dosing schedules for the CCR are as follows:

- Intensive Chemotherapy (IC): generally consists of cytarabine 100-200 mg/m^2 per day by continuous intravenous infusion for 7 days, plus three days of either daunorubicin 45-60 mg/m^2 per day or idarubicin 9-12 mg/m^2 per day for one cycle, followed by up to two consolidation cycles.[28] Our clinical expert commented that there are various different schedules for the administration of the IC treatment; however these differing combinations are unlikely to alter the clinical effectiveness of the drugs.
- LDAC: subcutaneously, 20 mg twice per day for 10 days.[28] This dosing practice was considered routine by our clinical expert.
- BSC: typically BSC will consist of prophylactic anti-infectious treatment (including fungal and antibiotic prophylaxis) and transfusion support (platelet, red blood cell and granulocyte transfusion).[28] This dosing practice was considered routine by our clinical expert.

Our clinical advisor (CR) commented that most patients are offered IC even if they ultimately receive an alternative treatment.

The ERG considers it a weakness to use a combined comparator in the decision problem, since it is possible that azacitidine could be effective and cost-effective (at a chosen cost-effectiveness threshold) versus some individual comparators but not others, but this would be obscured. This could result in either: azacitidine being used in patients when it would have been better for them and/or a better use of limited NHS resources for an alternative treatment to be used; or, azacitidine not being used in patients when it would have been better for the patient and/or a better use of limited NHS resources.

3.4 Outcomes

The outcomes in the company submission match those in the Scope. However, they have made the following amendments to progression free survival and response rate:

> *Progression free survival (PFS) - estimated from event-free survival (EFS) and relapse-free survival (RFS) for the purpose of economic modelling.*
>
> *Response rate – including complete remission, cytogenetic complete remission and partial remission.*

The company justify the alterations to PFS, by suggesting that PFS is not a standard endpoint for AML. Our clinical advisor (CR) believes that estimating PFS from EFS and RFS would be an appropriate alternative measure.

For the measure of response rate, the Scope specified response rate, including haematologic response and improvement, whereas the company reports that the outcome response rate included complete remission, cytogenetic complete remission and partial remission. These are only terminological differences.

3.5 Other relevant factors

In response to special considerations relating to equity and equality, the company make the following statement within their decision problem. The ERG has no comments on their statement.

> *AML presents primarily in the elderly population, with 64% of newly diagnosed cases in the UK in patients aged ≥65 years.[4] Equity of treatment of the elderly is a concern, as evident from a report published by the National Audit Office in January 2015.[38] AML is also an orphan disease.[39] The Cancer Patient Experience Survey in 2010 found that people with rarer forms of cancer reported a poorer experience of their treatment and care than people with more common forms of cancer.[40] Therefore, access where appropriate to a treatment such as azacitidine should help to promote equality for both elderly patients and those with rarer forms of cancer.*

4 Clinical effectiveness

4.1 Critique of the methods of review(s)

4.1.1 Searches

Celgene presented a literature search protocol to support its review of clinical effectiveness. This protocol included systematic searches of key biomedical databases using a literature search strategy, hand-searching of conference abstracts, and a search of ClinicalTrials.gov. The literature search was last updated in November 2015.

The bibliographic database searching used a search strategy that took the following form:

1. (controlled index terms for acute granulocytic, acute myeloblastic, acute myelomonocytic, acute monocytic or acute megakaryocytic leukemia or erythroleukemia or promyelocytic) *and*
2. (controlled index terms for Azacitidine, cytarabine, gemcitabine, deoxycytidine, etoposide, etopofos, fludarabine, anthracycline, mitoxantrone, daunorubicin or tioguanine or best supported care) *and*
3. (a range of search terms for study design (RCTs and observational), limits to remove studies conducted on animals and studies published in languages other than English. Systematic reviews and meta-analysis were also excluded from identification. The searches were date limited 2000 to current).

The search strategy was applied in the following bibliographic databases: MEDLINE (OVID), EMBASE (OVID) and Cochrane CENTRAL (OVID).

The following conference proceedings were hand-searched: European Haematology Association (EHA) Annual Congress, American Society of Haematology (ASH) Annual Conference, and the British Society for Haematology (BSH) Annual Scientific Meeting between January 2013 and April 2015. Finally, ClinicalTrials.gov was searched for relevant, unpublished studies.

The ERG believes the literature searching for clinical effectiveness studies was poorly conducted and reported.

The literature search strategies provided rely entirely on controlled indexing search terms. Free-text search terms for acute myeloid leuk(a)emia or AML (as used in the submission of cost-effectiveness literature searching) have been omitted. This is inappropriate for two reasons:

1. The searches (as presented) will only return studies that have been indexed. Newly published studies (or additional trial outcome data), or relevant studies published ahead of print, will likely be excluded from these searches as, whilst studies are uploaded to bibliographic databases on receipt, there is a delay between a study being published and then indexed. This point affects the currency of the literature searches and introduces bias into the identification of studies.
2. Any studies that have been incorrectly indexed but are of relevance to the decision problem would be missed by these searches. The inclusion of free-text terms for

AML, for example, would address this risk. This point affects the sensitivity of the literature search, which is poor.

The search strategy excludes the identification of systematic reviews and meta-analyses. In clarification, the manufacturer stated that the bibliographies of systematic literature reviews, meta-analyses and other included studies were used as a search strategy to identify further studies. As systematic reviews and meta-analyses were excluded from bibliographic searching it is not clear how they were identified and therefore how this particular search was actually conducted. This affects the replicability of the literature searching.

The literature searching is difficult to validate and it has been poorly reported.

The literature search strategies for MEDLINE and EMBASE have been combined and reported as if one search. This makes it difficult to validate and repeat the literature search used in MEDLINE as the EMBASE search strategy has been used and presented as the base search. Furthermore, the combination, and practicable use of study design literature search filters, is poorly considered when balanced against the decision problem of the review. This affects the transparency of the literature searching.

In view of these points, the ERG has undertaken its own scoping searches to ensure that no phase III RCTs have been missed by this review. Whilst basic scoping searches have not identified any additional studies, the ERG has been unable to validate aspects of this search, and the overall quality of the approach to literature searching is sufficiently poor, that it raises questions if it is truly fit for purpose.

4.1.1.1 Adverse events

Celgene did not undertake separate literature searches to identify studies reporting adverse events. In their submission, Celgene stated that (Source: Celgene submission, Section 4.12, p. 85):

> *No further studies that report additional adverse events …that are of relevance to the decision problem are available*

In clarification, Celgene confirmed that no separate literature searches to identify studies reporting adverse events were undertaken, stating that their literature searches were not limited by adverse event outcomes. Whilst it is true that Celgene's literature searches were not limited by outcomes, Celgene's literature searches were limited by study design. It is therefore possible that studies reporting adverse events may have been missed.

4.1.2 Inclusion criteria

Celgene's inclusion criteria are given below (*Table 2*) with an additional column added to the right of the table, taken from the Scope[36] for reference and comparison. Comments about the differences in inclusion criteria are outlined below the table.

Table 2: Scope of the literature review: PICOS criteria for study inclusion

Criteria	From Celgene	From Scope
	Definition	
Population	Older adult AML patients[a] with peripheral blood or BM leukaemic myeloblasts >20%, who either: • Are newly diagnosed with AML • Have developed AML secondary to "preleukaemic" blood disorders such as MDS or myeloproliferative disease • Have developed AML secondary to exposure to leukaemogenic therapy or agents with primary malignancy in remission for at least 2 years	Adults with acute myeloid leukaemia with bone marrow blasts more than 30%

Interventions/ comparators	• Azacitidine 75 mg/m^2
	• LDAC (20 mg SC once or twice a day for 10-14 days)
	• Decitabine 20 mg/m^2
	• Other high dose chemotherapy:
	  ○ Combination of etoposide or fludarabine (plus granulocyte-colony stimulating factor aka "G-CSF") with cytarabine (preferred for patients with cardiac disease)
	  ○ 7+3: continuous IV infusion of cytarabine for 7 days followed by 3 days of IV anthracycline push
	  ○ Combination of IV mitoxantrone, etoposide IV, and cytarabine
	  ○ Combination of IV daunorubicin, cytarabine, and etoposide
	  ○ Combination of IV cytarabine, daunorubicin, and oral thioguanine
	  ○ Combination of IV cytarabine and daunorubicin: 3+10 for cycle 1 followed by DA 3+8 for cycle 2 (standard for UK)
	• Best supportive care[b]
		Intervention: Azacitidine
		
		Comparators:
		• Intensive chemotherapy with an anthracycline in combination with cytarabine
		• Non-intensive chemotherapy with low dose cytarabine
		• Best supportive care which may include blood product replacement, antibiotics, antifungals and intermittent low dose chemotherapy with hydroxycarbamide

Outcomes	Studies are eligible if at least one of the following outcomes are included:[c]
	• Efficacy outcomes
	  ○ Overall survival
	  ○ Event-free survival
	  ○ Progression-free survival
	  ○ Relapse-free survival
	  ○ Complete response
	• Safety outcomes
	  ○ Treatment-related mortality
	  ○ Hospitalisation due to AE
	  ○ Grade 3 or 4 haematologic AEs
	  ○ Discontinuations due to AEs
	  ○ Discontinuations due to reasons other than disease progression
		The outcome measures to be considered include:
		• Overall survival
		• Progression free survival
		• Time to disease progression
		• Response rates, including haematologic response and improvement
		• Blood-transfusion independence
		• Infections
		• Adverse effects of treatment
		• Health-related quality of life

Criteria	From Celgene	From Scope
	Definition	
Study Design	• Randomised controlled trials and comparative non-randomised studies (prospective and retrospective observational studies) • Studies must compare two unique treatment classes (e.g. IC vs. IC or dose-ranging studies not eligible)	
Other	• English language only • Published in or after the year 2000 (Selected on the advice of a panel of haematologists who advised that there would be limited evidence of relevance pre the year 2000).	

Key: AE, adverse event; AML, acute myeloid leukaemia; BM, bone marrow; CR, complete response; EFS, event-free survival; IC, intensive chemotherapy; IV, intravenous; LDAC, low-dose cytarabine; MDS, myelodysplastic syndrome; OS, overall survival; PFS, progression-free survival; RFS, relapse-free survival; SC, subcutaneous.

Notes: a, Note that although the primary population of interest is those 65 years of age and older, this criteria was relaxed (e.g. 55 years of age and older) to ensure sufficient evidence was available; b, It was expected that definitions in best supportive care would vary; c, Note that additional outcomes were of interest, but only those identified in the table above were used to guide the selection of studies; safety outcomes were extracted only for those studies providing efficacy data.

Source: Celgene submission, Table 7, pp. 39–40 and NICE Scope[36]

4.1.2.1 Population

Celgene's population was broader (bone marrow blasts more than 20%) than the Scope (bone marrow blasts more than 30%).[36] The company explained that this was to capture studies on bone marrow blasts over 20% that had included sub-analysis on the 30% bone marrow blast population.

Conversely, Celgene's population was narrower (older adult AML patients) than the Scope (adults with AML).[36] Older adults were defined by Celgene for their systematic review as a population over 55 years. In order to check if any studies were excluded based on population; the ERG checked the reasons given in table of excluded studies (Celgene appendices to submission, Appendix 2, Table 1, pp. 10–11) and found none. However, we cannot exclude the possibility studies may have been excluded from title/abstract screening.

4.1.2.2 Interventions/comparators

The intervention/comparators for inclusion broadly match the Scope.[36] However, there are some key differences. For the intervention azacitidine and two of the comparators (LDAC and decitabine) drug doses were included. It is not clear whether studies were excluded based on the drug administration dose, nevertheless drug administration doses are not specified within the Scope.

The comparators for inclusion reported by the company include all of those reported in the Scope and also decitabine.[36] The company also include specific drug combinations to be given alongside intensive chemotherapy rather than, 'with an anthracycline in combination with cytarabine' as per the Scope. This may have resulted in relevant studies being excluded.

4.1.2.3 Outcomes

The outcomes overall survival, progression-free survival match the Scope.[36] However, event-free survival, relapse-free survival and complete response are not within the Scope. Other outcomes within the Scope, but not in the inclusion table submitted by the company include: time to disease progression; response rates, including haematologic response and improvement; blood transfusion independence; infections; adverse effects of treatment and health related quality of life. In order to check if any studies were excluded based on outcomes; the ERG checked the reasons given in table of excluded studies (Celgene appendices to submission, Appendix 2, Table 1, p. 10–11) and found none. However, we cannot exclude the possibility studies may have been excluded from title/abstract screening.

4.1.2.4 Study Design

The Scope did not restrict study design. However, the NICE reference case guide to the methods of technology appraisal 2013 (Chapter 5.2.3)[41] recommends studies should be restricted to RCTs and when they are not available, non RCTs. Studies included in the company submission were RCTs, observational studies and information from registries. We are satisfied the study designs meet the reference case.

4.1.2.5 Other

Celgene applied an English language restriction to their systematic review. Language was not given as a reason for exclusion from full-text screening (Celgene appendices to submission, Appendix 2, Table 1, p. 10–11). However, we cannot exclude the possibility studies may have been excluded from title/abstract screening.

Another restriction was to exclude studies published before 2000. The decision to restrict the studies to only include studies post-2000 was (Source: Celgene submission, Section 4.1.2, p. 40):

> *Selected on the advice of a panel of haematologists who advised that there would be limited evidence of relevance pre the year 2000*

The ERG notes that azacitidine was used for primarily haematologic malignancies in the 1960s-1980s.[42-44] Consequently, studies excluded based on date limitation could add relevant information for the use of azacitidine in treating AML. Reference citations for the 15 RCT studies reported as excluded based on publication date (Celgene appendices to submission, Appendix 2, Table 1, p. 10–11) were requested and subsequently examined. None of the 15 studies would have been eligible for inclusion in this submission.

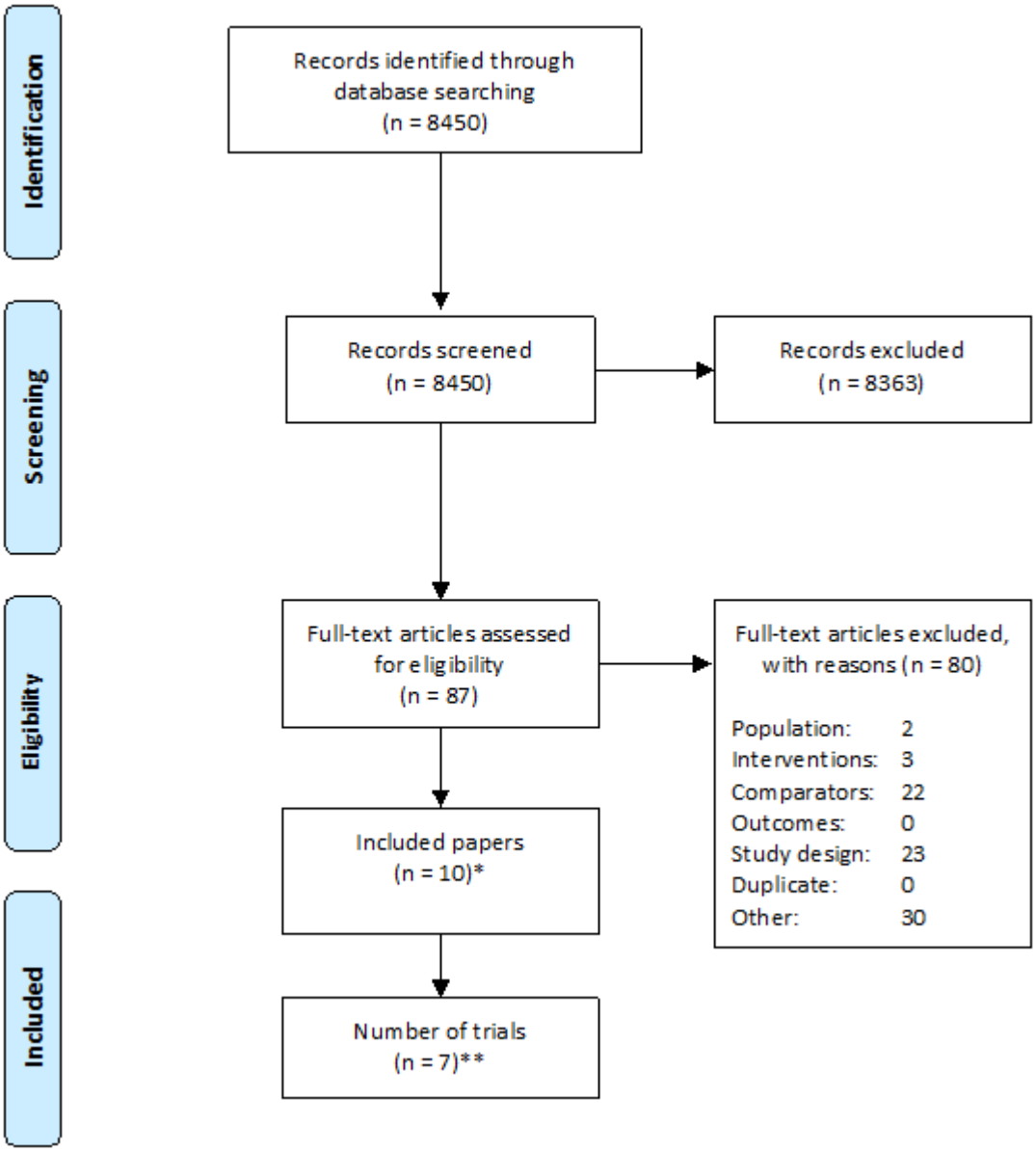
4.1.2.6 Study selection

Celgene's submission explains the process used in study selection (i.e., that two researchers independently reviewed the abstracts and the full-texts of studies, that discrepancies between investigators were resolved by involving a third investigator and coming to a consensus). These are standard procedures for systematic reviews.[41]

From the 8,450 citations the company identified from their searches, 8,363 citations were excluded and 87 were taken to full-text screening at the abstract screening stage. From the full-text screening, 80 citations were excluded with reasons for exclusion provided (Celgene appendices to submission, Appendix 2, Table 1, pp. 10–11). The company go on to explain

that three further materials were added from manual searches of literature databases and conference proceedings. One of the three additional materials identified from the manual searches was the pivotal trial AZA-AML-001. The company included ten citations from seven trials. The PRISMA diagram reported in Celgene's submission is copied below (*Figure 1*).

Figure 1: PRISMA study flow diagram



Notes: * 3 materials added; ** 4 RCTs and 3 observational studies.
Source: Celgene submission, Figure 4, p. 42

From the seven included studies, the company go on to exclude a further five studies. The exclusion of these five studies is justified by the company on the basis that their inclusion criteria were over-inclusive. Four studies were excluded as the population was those with bone marrow blasts over 20% and not 30% as per the Scope. The fifth study was excluded as the comparator treatment was decitabine, a comparator not included in the decision problem.

The remaining two studies were an RCT by Dombret et al. 2015.[45] and an observational study by Lao et al. 2015.[46] The main focus of the submission was the RCT study by Dombret et al. (AZA-AML-001). The company also refer to registry data from three countries (Austria, Spain and France) within their submission.

4.1.3 Data extraction

The submission explains the process of data extraction used, which is in line with the standard review process (Source: Celgene submission, Section 4.1.2, p. 40)

> *Two investigators independently extracted data on study characteristics, interventions, patient characteristics at baseline, and outcomes for the study populations of interest for the final list of selected eligible studies. Any discrepancies found between the data extracted by the two data extractors were resolved by involving a third reviewer and coming to a consensus.*

The ERG notes that in relation to the pivotal RCT by Dombret et al.,[45] all the typical data (including but not limited to: participant inclusion criteria, baseline characteristics, methods, primary and secondary study outcomes and adverse events) has been extracted from the paper, but that there were several typographic errors in the extracted data reported in their submission. The ERG has referred to the original publication and the clinical study report for AZA-AML-001[37] to resolve discrepancies where they have been identified.

4.1.4 Quality assessment

Details of the company's critical appraisal of the RCT AZA-AML-001,[45] alongside our critique, can be seen below in *Table 3*. The critical appraisal appears (since no reference is given to the tool used) to have been performed by the company using the CRD assessment criteria for risk of bias in RCTs. However, they have slightly adjusted the wording of the questions, as they do not match exactly. The meanings behind the questions remain the same.

Table 3: Critical appraisal of AZA-AML-001

Critical appraisal criterion	Celgene's Assessment	ERG Comment
Was randomisation carried out appropriately?	Yes Patients were randomised in a 1:1 ratio. Randomisation was performed using an IVRS. Patients were stratified at randomisation by: • CCR selection (IC, LDAC or BSC), • ECOG performance status at baseline (0–1 versus 2) • Cytogenetics (intermediate-risk versus poor-risk)	Further information on the randomisation process from the paper reports that: a central, stratified, and permuted block randomisation method and IVRS were used to randomly assign 1:1 to received azacitidine or CCR. This is an acceptable system for randomisation.
Was the concealment of treatment allocation adequate?	Open-label study. Blinding of study treatment was not feasible due to multiple comparators and routes of administration. However, all central reviewers were blinded to subject treatment assignment.	Celgene have not answered this question. Allocation sequence concealment obtains strict implementation of an allocation sequence without prior knowledge of the intervention assignments. Methods for allocation concealment refer to techniques used to implement the sequence, not to generate it. For this trial, central randomisation ensured allocation sequencing was adequately concealed.
Were the groups similar at the outset of the study in terms of prognostic factors?	Yes. Patient demographics in the azacitidine and combined CCR groups were well balanced in terms of age, age distribution, sex, geographic location, race, weight and BSA. The azacitidine and combined CCR groups were also comparable for all baseline disease characteristics (including AML classification, prior history of MDS, time since AML diagnosis, ECOG performance status and cytogenetic status), with the exception of prior anticancer systemic therapies.	As Celgene have written, demographics between azacitidine and combine CCR are well balanced and we are satisfied with this assessment. More meaningfully perhaps, baseline characteristics for the three individual CCR arms could have been compared to the azacitidine arm split by the CCR assignment prior to randomisation. However, this data was not available. Confusingly, Celgene contradict themselves by reporting (p. 26) that they provided evidence to the CHMP that outcome failures were due to an imbalance of patients' baseline characteristics/prognostic factors.
Were the care providers, participants and outcome assessors blind to treatment allocation?	Although the trial was open-label, all central reviewers were blinded to subject treatment assignment. Evaluations by central review were used for the statistical efficacy analyses. The independent review committee which reviewed and confirmed the haematologic responses and durations was blinded to treatment, investigative site, and subject identifier.	Since the study was open labelled, the care providers and participants could not be blinded to treatment allocation. Awareness of treatment allocation will have introduced the potential for bias within the study, particularly with reporting of adverse events. All central reviewers were however, blinded to treatment assignment.

Critical appraisal criterion	Celgene's Assessment	ERG Comment
Were there any unexpected imbalances in drop-outs between groups?	No. The most common reasons for discontinuation from the treatment phase in both the azacitidine and CCR groups were occurrence of an AE (36.9% and 26.7%, respectively) or death (22.0% and 23.5%, respectively). Discontinuations due to occurrence of an AE or study closure were more common in the azacitidine group whereas discontinuations due to withdrawal of consent were more common in the CCR group, with the highest percentage in the BSC group. The percentages of subjects who were discontinued from treatment because of death or disease progression were comparable between the azacitidine and the CCR treatment groups. No subject discontinued due to loss of follow-up or protocol violation in the azacitidine group. One subject discontinued due to loss of follow-up in the IC group and one subject discontinued due to protocol violation in the LDAC treatment group.	The drop-outs provided by Celgene represent the figures reported in the RCT. However, neither Celgene nor the RCT report the drop outs for the three CCR treatments separately, therefore it is unknown whether the actual treatments that make up the CCR are comparatively different to azacitidine. The dropout rates due to an AE look imbalanced for azacitidine (36.9%) compared to the CCR group (26.7%).
Is there any evidence to suggest that the authors measured more outcomes than they reported?	No. All treatment outcomes were reported other than those that are not currently available for analysis (exploratory molecular markers).	The ERG agrees with Celgene's response to this question.
Did the analysis include an intention-to-treat analysis? If so, was this appropriate and were appropriate methods used to account for missing data?	The ITT population was used for the analysis of the primary and secondary efficacy endpoints. The ITT population was the most appropriate population as it included all randomised patients.	Yes – the main analysis adopts 'intention to treat' principles.

Key: AE, adverse event; AML, acute myeloid leukaemia; BSA, body surface area; BSC, best supportive care; CHMP, Committee for Medicinal Products for Human Use; CCR, conventional care regimens; ERG; evidence review group; IC, intensive chemotherapy; ITT, intent-to-treat; IVRS, interactive voice response system; LDAC, low-dose cytarabine; MDS, myelodysplastic syndromes. Clinical effectiveness results of the relevant randomised controlled trials.

Source: Celgene submission, Table 17, p. 64

4.1.5 Evidence synthesis

From the searches, one RCT was identified. Therefore synthesis of the evidence was not required.

4.2 Critique of the trials of the technology of interest, their analysis and interpretation

4.2.1 Methods

The single RCT (study name AZA-AML-001; main publication by Dombret et al. 2015[45]) identified was presented in detail within the submission.

4.2.1.1 Study objectives

The company submission reports the study objectives as follows (Source: Celgene submission, Section 4.3.1, p. 43):

> *The primary objective of the study was to demonstrate superiority in OS of azacitidine compared with combined CCRs in subjects aged 65 years or over who had newly diagnosed AML with more than 30% BM blasts according to the WHO criteria,[14, 47] and who were not eligible for haematopoietic stem cell transplantation. Overall survival was defined as time from randomisation to death from any cause.*

Secondary objectives included 1-year OS rate, EFS, RFS, overall remission rate, cytogenetic complete remission (CRc) rate, safety and toxicity assessments, HRQoL and health resource utilisation.

The primary objective concurs with the primary outcome; an outcome specified within the NICE Scope.[36] The secondary objectives correspond in parts to the outcomes from the Scope. The differences are the same as those already discussed in *Section 4.1.2.3 (page 32)*.

4.2.1.2 Study design and treatment

The study AZA-AML-001 was a multicentre (over 18 countries), randomised, open-label, parallel-group study.

Before randomisation, the most appropriate CCR (IC, LDAC or BSC) was determined by investigators on the basis of age, Eastern Cooperative Oncology Group Performance Status (ECOG PS), comorbidities and regional guidelines and/or institutional practice. Once the CCR had been chosen, a central, stratified, and permuted block randomisation method and interactive voice response system was used to randomly assign 1:1 to receive azacitidine with BSC or the pre-selected CCR. The random treatment assignment was concealed so that investigators and subjects did not know in advance the next treatment assignment. Following randomisation and drug administration, follow-up appointments were scheduled once per week during the first two treatment cycles, then 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 4*.

By design, there is awareness of the treatment allocated for both the patient and primary care givers from an open-labelled study. Awareness of treatment allocation will have introduced the potential for bias within the study, particularly with reporting of adverse events. Based on the treatments administered within the study, an open-label study design was the most appropriate study design to be utilised. Treatment protocols for administration were confirmed as appropriate by our clinical advisor (CR) and follow-ups were similar and appropriate in time.

Table 4: Treatment protocol

Treatment	Administration	Data collection[a]
Azacitidine & Best Supportive Care	75 mg/m^2/day, subcutaneously for 7 consecutive days per 28-day treatment cycle for at least 6 cycles. Dosing could be reduced or delayed as needed until the blood count recovered. BSC: as for BSC and including transient use of hydroxyurea (hydroxyurea was not allowed within 72 hours before or after azacitidine administration).	Within 7 days before initiation of every second cycle beginning at cycle 3
Best supportive care	Included, but was not limited to, treatment with red blood cell or whole blood transfusions, fresh frozen plasma transfusions, platelet transfusions, antibiotic and/or antifungal therapy, and nutritional support. Hydroxyurea use was permitted under certain conditions	On day 1 of every third cycle (a BSC cycle was defined as 28 days), beginning at cycle 4.
LDAC and BSC	20 mg of cytarabine twice per day, subcutaneously for 10 days per 28-day treatment cycle for at least 4 cycles). Dosing could be reduced or delayed as needed until the blood count recovered. BSC: as for BSC and including transient use of hydroxyurea	Within 7 days before initiation of every second cycle beginning at cycle 3
IC and BSC	Cytarabine 100-200 mg/m^2/day by continuous intravenous infusion for 7 days, plus on days 1-3 if cytarabine an anthracycline (either daunorubicin 45-60 mg/m^2/day or idarubicin 9-12 mg/m^2/day) for 1 cycle. Followed by up to 2 consolidation cycles (i.e., the same anthracycline regimen as used at induction and the same cytarabine dose used for induction but administered for 3 to 7 days) for those achieving complete response or partial response. Re-induction was not allowed. BSC: as for BSC and including transient use of hydroxyurea	At screening and within 7 days before each treatment cycle

Key: BSC, best supportive care; IC, intensive chemotherapy; LDAC, low-dose cytarabine
Notes: a, consisting of cytogenetic testing and pathological samples (BM aspirates, BM biopsies, and peripheral blood smears) to confirm diagnosis.

Concomitant medications were kept to a minimum, but where necessary and where unlikely to interfere with trial drugs, were given at the discretion of the investigator. Concomitant medications included and excluded were as described in *Table 5*.

Table 5: Concomitant medications in/excluded

Included	Excluded
• Hydroxyurea • Serotonin • Blood product support (red blood cells and platelets) • Myeloid growth factors (only for the treatment of neutropenic infections, prophylactically during IC treatment, or in subjects with two or more previous episodes of neutropenic infection who were at risk of subsequent neutropenic infection. For subjects who developed an absolute neutrophil count (ANC) <0.5 x 10^9/L, administration of prophylactic fluoroquinolone was permitted.) • Erythropoietic agent.	• Clofarabine • Decitabine • Targeted agents (e.g. FLT-3 antagonists) • Systemic anticancer therapy (excluded hydroxyurea) • Oral retinoids (topical retinoids were permitted) • Use of any other investigation drug or therapy

Key: ANC, absolute neutrophil count; FLT-3, FMS-like tyrosine kinase 3

Subjects had follow-up visit for the collection of AEs 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, subjects were followed for survival on a monthly basis until death, lost to follow-up, withdrawal of consent, or end of the study. Considering the short survival duration for people with AML, perhaps assessing AEs on a monthly basis was not frequent enough to capture all the changes over time for AEs.

4.2.1.3 Study duration

Celgene report the following for the planned study duration (Source: Celgene submission, Section 4.3.2.1, p. 46):

> *The expected duration of the study was 31 months. This time frame consisted of a 19-month subject enrolment period, followed by 12 months of subject treatment and observation. The study was planned to conclude 12 months after the last subject was randomised.*

Study duration was suitable, enabling adequate assessment of the outcomes following treatment for AML.

4.2.1.4 Blinding

The treatment of AML within AZA-AML-001 necessitated an open-labelled design due differing routes of administration (subcutaneous injection / intravenous infusion) and time periods of treatment. Open-label design creates an opportunity for bias, particularly for reporting of AEs. Central review of peripheral blood, BM samples and cytogenetics was conducted by a pathologist and cytogeneticist blinded to treatment. AML classification for each person was determined by local investigators at study entry. The Independent Review Committee which reviewed and confirmed the International Working Group responses and durations was blinded to treatment, investigative site, and subject identifier. Blinding of the central reviewers was appropriate.

4.2.1.5 Inclusion/exclusion

Table 6 gives the inclusion/exclusion criteria from the trial.[45] Critique of these follows the table.

Table 6: Eligibility Criteria

Key inclusion criteria	Key exclusion criteria
• Newly diagnosed, histologically confirmed de novo or secondary AML • BM blasts >30% • Adults aged ≥65 years • Not considered eligible for hematopoietic stem cell transplantation, • Intermediate- or poor-risk cytogenetics (NCCN 2009 criteria) • ECOG performance status of ≤2 • White blood cell count ≤15 × 10^9/L	• Acute promyelocytic leukaemia with t(15;17)(q22;q12) • AML with inv(16)(p13.1q22) or t(16;16)(p13.1;q22), t(8;21)(q22;q22), or t(9;22)(q34;q11.2) • AML arising from previous hematologic disorders other than myelodysplastic syndrome (e.g., myeloproliferative neoplasms) • Other malignancies • Uncontrolled systemic infection • Prior recipient of decitabine, azacitidine, or cytarabine treatment • Prior AML therapy (except hydroxyurea, which was allowed up to 2 weeks before the screening haematology sample was taken) • Any experimental drug within 4 weeks of starting study treatment

Key: AML, acute myeloid leukaemia; BM, bone marrow; ECOG, Eastern Cooperative Oncology Group
Source: Dombret et al. 2015[45]

Additional exclusion criteria provided in the company submission but not specified in the RCT paper include: Prior BM or stem cell transplantation, proven central nervous system leukaemia, inaspirable BM, unstable angina, significant cardiac arrhythmia, or New York Heart Association class 3 or 4 chronic heart failure, pregnant or lactating women, active viral infection with known HIV or viral hepatitis B or C, known or suspected hypersensitivity to azacitidine or mannitol, use of any experimental drug or therapy within 28 days prior to day 1 of cycle 1, unwilling or unable to complete PRO assessments without assistance or minimal assistance, any condition, including laboratory abnormalities, which would place the subject at an unacceptable risk, any significant medical condition, including the presence of laboratory abnormalities, or psychiatric illness which would interfere with subject participation, any condition that confounded the ability to interpret data from the study.

Of note, the exclusion criteria related to cytogenetics are all primarily *favourable* characteristics (in line with the inclusion criteria of intermediate- or poor-risk cytogenetics), except for t(9;22) which is classed as poor-risk and is managed as the blast crisis phase in chronic myeloid leukaemia with the addition of tyrosine kinase inhibitors.

4.2.1.6 Location

The location of investigation sites were reported to be as follows (Source: Celgene submission, Section 4.3.4, p. 48):

> *Screening was conducted in 107 investigational sites, of which, 98 sites randomised at least one patient across 18 countries in different geographic regions. Locations included: Asia (12 sites); Australia (6 sites); the US/Canada (12 sites); Eastern Europe (12 sites); and Western Europe and Israel (56 sites). These included 5 sites in the UK which in total randomised 26 patients: Oxford (n=4), Bournemouth (n=1), St Bartholomew's (n=13), King's College (n=4) and Wolverhampton (n=4).*

In terms of the 26 people recruited from the UK, these people made up 5.3% of the total patient population from the trial. Fifteen UK people we randomised to receive azacitidine and 11 people received CCR (three to BSC and eight to LDAC). People recruited from Western Europe/Israel made up 48.8% of the total patient population.

4.2.1.7 Study endpoints

The study endpoints and definitions are presented in *Table 7*.

Table 7: Study endpoints

End point	Definition
Primary end point	
Overall survival (OS)	Defined as the time from randomisation to death as a result of any cause
Secondary end points	
1-year OS rate	No definition provided in either submission or Dombret et al 2015. Presumed to mean the number (or percentage) of people still alive 1 year post randomisation.
Event free survival (EFS; not in Scope)	Defined as the interval from the date of randomisation to the date of treatment failure,[a] progressive disease, relapse after complete response (CR) or complete response with incomplete blood count recovery (CRi), death from any cause, or loss to follow-up, whichever occurred first
Relapse free survival (RFS; not in Scope)	Defined only for subjects who achieved CR or CRi and was measured as the interval from the date of first documented CR or CRi to the date of relapse, death from any cause, or loss to follow-up, whichever occurred first. Relapse defined as either the recurrence of >5% blasts in the peripheral blood following Cr or CRi, (the percentage of peripheral blood blasts must have been ≤5% at the time of CR or CRi or a single finding of >15% blasts in the BM following a CR or CRi
Overall remission rate (CR+CRi)	Conditions for CR include: the BM should contain fewer than 5% blast cells; ANC ≥1,000/μL and Platelet count ≥100,000/μL. No RBC, platelet, or whole blood transfusions for 1-week prior to the haematology assessment used for the response evaluation. CRi was defined as a morphologic complete remission but the ANC count may be <1,000/μL and/or the platelet count may be <100,000/μL.
Duration of remission (CR + CRi)	Defined as the time from the date of CR or CRi until the date of relapse from Cr or CRi
Cytogenetic complete remission rate (CRc)	Defined as morphologic CR with a return to a normal karyotype at the time of CR (based on ≥10 metaphases)

End point	Definition
Progressive disease (PD)	Defined as either: 1) a >50% increase in BM blast count from baseline that persists for at least 2 BM assessments separated by at least 1 month, or if the baseline BM blast count is >70% and persists for 2 post-baseline BM assessments separated by at least 1 month, or 2) a doubling of the baseline absolute peripheral blood blast count that persists for at least 7 days and the final absolute peripheral blood blast count is >10 x 10^9/L. The date of PD is defined as the first date that there was either a >50% increase in BM blast count from baseline, a persistence of BM blasts >70% in subjects with a baseline BM blast count of >70%, or a doubling of the peripheral blood blast count.
Partial remission (PR)	Defined as an ANC ≥1,000/μL and platelet count ≥100,000/μL with a >50% decrease in the percentage of BM blasts to 5–25%
Stable disease (not in Scope)	Defined as any evaluable time point where criteria for all other response categories (i.e., CR, CRi, PR, progressive disease, treatment failure, not assessable) are not met
Safety/tolerability	Covering type, frequency, severity, and relationship of AEs to study treatments; physical examinations, vital signs; clinical laboratory evaluations; and concomitant medication/therapy
Patient-reported quality of life	Using the European organisation for research and treatment on cancer, quality of life questionnaire C-30 (EORTC QLQ-C30) Completed on day 1 of cycle 1 (baseline), every other cycle thereafter, and at the end-of-study visit
Measures of healthcare resource utilisation (HCRU)	Any consumption of healthcare resources directly or indirectly related to the treatment of the subject. Five items of HCRU were collected: inpatient hospitalisations, transfusions, procedures or surgeries, and concomitant medications
Additional endpoints	
Transfusion status (RBC and platelet transfusion status [dependence or independence])	On-treatment RBC/platelet transfusion independence was defined as the absence of any RBC/platelet transfusions for 28 or 56 consecutive days during the treatment period
Peripheral blood counts	To include platelets, absolute neutrophil count, haemoglobin, white blood cell, and blasts

Key: AE, adverse event; AML, acute myeloid leukaemia; ANC, absolute neutrophil count; BM, bone marrow; CR, complete response; CRi, complete response with incomplete blood count recovery; DNA, deoxyribonucleic acid; EFS, event-free survival; EORTC, European Organization for Research and Treatment on Cancer; Hgb, haemoglobin; IWG, international Working Group; MDS, myelodysplastic syndrome; PR, partial remission; QLQ, quality of life questionnaire; RBC, red blood cell; RFS, relapse-free survival; WBC, white blood cell.

Notes: a, Treatment failure defined as death during cycle 1 or within 28 days of the last dose and prior to day 1 of cycle 2

These endpoints are common and reasonable for a study investigating AML. A similar AML study[48] investigating a different drug to that of AZA-AML-001, report fewer endpoints than AZA-AML-001. However, the endpoints that they do report, match those of AZA-AML-001.

4.2.1.8 Statistical methods

4.2.1.8.1 Analysis population

The different populations reported within Celgene's submission for their analyses, along with their definitions are presented in *Table 8*. Celgene also included analysis of a modified ITT population and an evaluable population. These analyses were deemed not relevant to the question asked from this STA by the ERG.

Table 8: Analysis Population

Analysis Population	Definition
Intent-to-treat population (ITT)	All subjects who were randomised, independent of whether or not they received study treatment. The ITT population was used for the analysis of the primary and secondary efficacy endpoints. Subjects in the ITT population were analysed as randomised.
HRQoL evaluable population	All randomised subjects who completed the baseline HRQoL assessment (day 1) and had at least one follow-up assessment.
Safety population	All randomised subjects who had received at least one dose of trial drug and had at least one post-dose safety assessment. Subjects who were randomised to BSC within the CCR group were considered to be included in the safety population is that had at least one post-randomised safety assessment. Drug exposure and all safety analyses were based on the safety population. All subjects were analysed according to the initial treatment they received.

Key: BSC, best supportive care; CCR, conventional care regimen; ITT, intention-to-treat; HRQoL, health-related quality of life

The ITT population is the most appropriate population to use for analysis and the definition of the ITT population is correct. There is a risk of bias for the HRQoL population, as this population represents those who were well enough to complete the questionnaire and provide data. The safety population is defined appropriately.

4.2.1.8.2 Determination of sample size

Celgene report in their submission the determination of sample size to have been as follows (Source: Celgene submission, Section 4.4.3., p. 54):

> *The equality of OS curves was to be compared between the azacitidine and combined CCR groups using a stratified log-rank test. The planned sample size was approximately 480 subjects (240 per treatment arm), calculated on the assumption of a median OS of 10.5 months in the azacitidine arm and 7.5 months in the combined CCR arm (40% improvement), with a dropout rate of 1% from both treatment groups. The investigator selection of CCR was anticipated to be 50%, 30%, and 20% of subjects to the IC, LDAC, and BSC groups, respectively. This design required 374 deaths to allow the demonstration of a statistically significant difference in OS at a one-sided significance level of 0.025 with at least 90% power to detect a constant HR of 0.71.*

The study was powered for azacitidine compared to combined CCR. The results would have been more meaningful if the study had been powered to each of the CCR treatments

individually. Celgene anticipated the selection of CCR to be 50:30:20 for IC:LDAC:BSC. The actual study recruitment to CCR has the ratio 18:64:18. Celgene were asked to comment on the difference in anticipated selection of CCR and actual selection on CCR. Their anticipated selection for CCR was based on an educated guess since there was little real-world data to inform the prospective split.

4.2.1.8.3 Primary and secondary efficacy analysis

The company report the following for their primary efficacy analysis (Source: Celgene submission, Section 4.4.4, p. 54):

> *The primary efficacy analysis was performed using the ITT population. The analysis of the primary efficacy endpoint was conducted using an unstratified log-rank [test] and a stratified log-rank test (stratified by CCR selection, ECOG performance status, and cytogenetic risk status). The Kaplan Meier (KM) method was used to estimate the survival distribution functions for each treatment group. KM estimates for median OS, 25th and 75th percentiles, and associated two-sided 95% CIs were summarised for each treatment group (both unadjusted for the stratification variables and within strata). Additionally, the numerical difference and associated 95% CI in the median, and the 25th and 75th percentiles between the two treatment groups (azacitidine vs. CCR) were presented for the unstratified KM estimates.*
>
> *Cox proportional hazards models (unstratified and stratified) were used to estimate the hazard rate ratio and the corresponding 95% CI for azacitidine vs CCR.*
>
> *Surviving subjects were censored upon study discontinuation (loss to follow-up, withdrawal of consent) or at the end of the post-study follow-up.*

The company report the following for their secondary efficacy analysis (Source: Celgene submission, Section 4.4.5, p. 55):

> *All secondary endpoints were analysed using the ITT population, except for HRQoL and healthcare resource utilisation (HCRU). Analyses for both HRQoL and HCRU were conducted using a HRQoL evaluable population, defined as all randomised subjects who completed the baseline HRQoL assessment (day 1) and had at least one follow-up assessment.*
>
> *Kaplan-Meier methods were used to estimate the 1-year survival probabilities for time to death from any cause and death probabilities at 30 and 60 days.*
>
> *Time-to-event endpoints (EFS and RFS) were analysed using the same methods as the primary efficacy analysis, but without stratification. For EFS, subjects who were alive and event-free were censored at the date of their last response assessment, and for RFS, subjects who were in continuous CR or CRi were censored at the date of their last response assessment.*
>
> *Haematologic status was explored by examining the percentage of responders, defined as CR and CRi, and the duration of remission, CRc, peripheral blood counts, and transfusion requirements. All responses were based on the modified International Working Group (IWG) response criteria for AML.*

> *For duration of remission, subjects who were lost to follow-up or were alive at follow-up without documented relapse were censored at the date of their last response assessment. Summary statistics included KM estimates of median duration of remission, and 1-year cumulative incidence of relapse for each treatment group.*
>
> *For transfusion status, subjects who maintained red blood cell/platelet transfusion independence to the end of the treatment period were censored at the date of treatment discontinuation or death, whichever was sooner. Duration of transfusion independence was estimated and summarised using KM methods.*
>
> *For HRQoL analyses, the mean change from baseline for each domain at each time point was compared with the minimal important difference to determine whether the change was clinically meaningful. A mean change of at least 10 points on the standardised domain scores was required to be considered meaningful.*[49]
>
> *All reported log-rank or Fisher's exact test p values for secondary endpoints are nominal.*

The majority of the statistical methods used in the trial to analyse time-to-event data (and in particular, the primary efficacy outcome, overall survival) assume proportional hazards or have reduced efficiency in the presence of non-proportional hazards. There is no justification given for expecting this assumption to hold, and considering the results (e.g., *Figure 2, page 47*) this assumption was not reasonable.

In the presence of non-proportional hazards, appropriate alternatives to the log-rank test employed in the trial would be the Wilcoxon–Breslow–Gehan and Peto–Peto–Prentice tests (the choice between these based on the assessment of there being any differences in censoring patterns). In the presence of non-proportional hazards, there is no simple alternative to using a hazard ratio, but statistics such as the difference in restricted mean survival could be meaningful.

Using both stratified and unstratified tests is not directly justified. Overstratification can lead to loss of information, but unstratified tests are not appropriate when there is heterogeneity between strata.[50] Given the variables used for stratification are considered prognostic indicators, this suggests that the stratified analyses are more appropriate. Furthermore, the set of variables used for stratification was reduced in the event that stratification with the full set of variables would lead to individual strata with fewer than 16 patients, which should have reduced the risk of overstratification.

The ERG considers the censoring events loss to follow-up and withdrawal of consent may be informative for overall survival (which would violate the necessary assumptions for Kaplan–Meier analyses) but the most common reason for censoring was the patient being alive at the time of study closure,[37] so this is unlikely to have significantly impacted on results.

4.2.2 Results

4.2.2.1 Population distribution

In total, 488 people were randomised. Of these, 241 subjects were randomised to receive azacitidine, and 247 people were randomised to receive conventional care treatment. The

number of participants evaluable for each of the different population (ITT, safety, evaluable, HRQoL evaluable), are presented in *Table 9*. *Table 9* also presents the distribution of pre-selected CCR within the 247 people randomised to the azacitidine arm.

Table 9: Population distribution for analysis

Analysis population	Azacitidine[a]				CCR			
	BSC only (N=44)	LDAC (N=154)	IC (N=43)	Total (N=241)	BSC only (N=45)	LDAC (N=158)	IC (N=44)	Total (N=247)
ITT	44	154	43	241	45	158	44	247
Safety	42	151	43	236	40	153	42	235
Evaluable	35	114	30	179	25	132	34	191
HRQoL evaluable	—	—	—	157	—	—	—	134

Key: AZA, azacitidine; BSC, best supportive care; CCR, conventional care regimens; HRQoL, health related quality of life; IC, intensive chemotherapy; ITT, intent-to-treat; LDAC, low-dose cytarabine.
Notes: a, Number of patients randomised to azacitidine for each prespecified CCR
Source: Celgene submission, Table 14, p. 59

4.2.2.2 Baseline characteristics and demographics

Baseline characteristics of the ITT population are summarised in *Table A1* (*Appendix 1*) and baseline disease characteristics are presented in *Table A2* (*Appendix 1*). The demographic characteristics are well balanced between those randomised to azacitidine and the combined CCR. Those in the IC group were slightly younger than any of the other treatment groups. Conversely, those in the BSC group were slightly older than other treatment groups. These age differences are to be expected based on the participant demographic typically assigned to these types of CCR treatments. The azacitidine and combined CCR treatment groups were comparable for baseline disease characteristics.

4.2.2.3 Treatment exposure

Median treatment cycles were as follows:

- Azacitidine treatment cycles received 6 (range, 1-28 cycles);
- IC treatment cycles received 2 (range, 1-3 cycles);
- LDAC treatment cycles received 4 (range, 1-25 cycles);
- BSC treatment cycles received 65 days (range, 6-535 days).

From the azacitidine group, 52.5% received 6 or more treatment cycles and 32.2% received 12 or more treatment cycles. From the LDAC group, 35.9% received 6 or more treatment cycles and 17.6% received 12 or more treatment cycles. Cumulative patient-years of study drug exposure were 174.9 for azacitidine, 82.9 for LDAC, 14.1 for IC, and 9.6 (i.e., time on study) for BSC.[45]

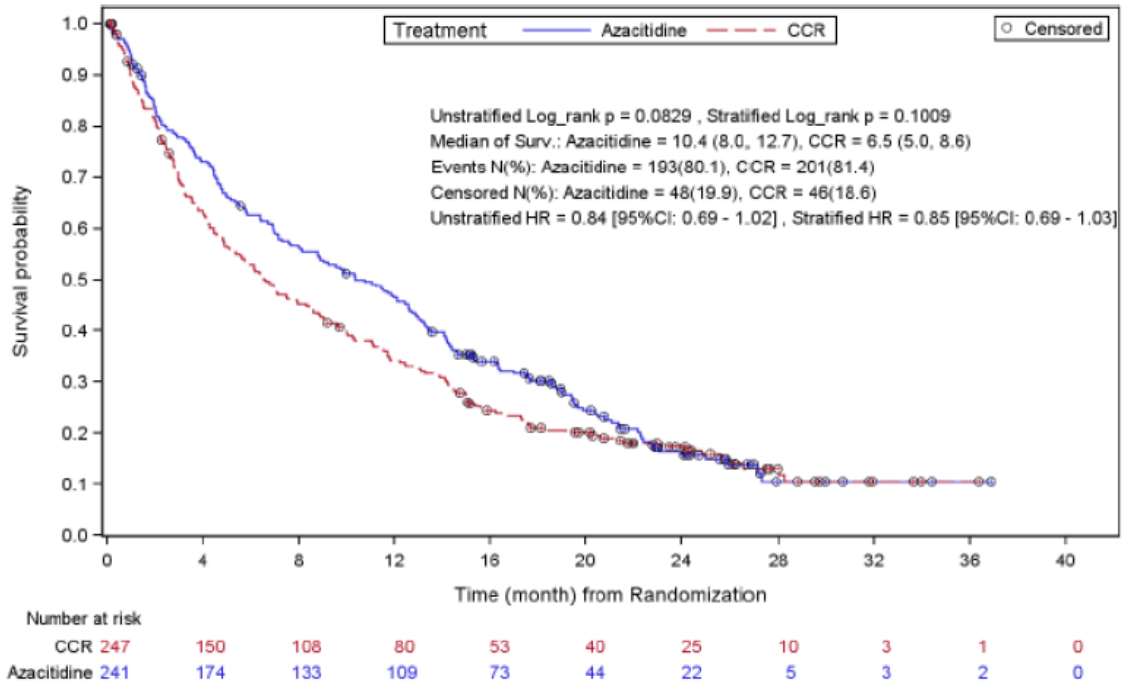
4.2.2.4 Clinical effectiveness results

4.2.2.4.1 Primary efficacy analysis – overall survival

The median duration of follow up was 24.4 months. By study end, 394 deaths (80.7%) had occurred; 193 (80.1%) in the azacitidine group and 201 (81.4%) in the CCR group. The Kaplan-Meier plot of time to death from any cause is presented in *Figure 2* and a summary of OS is presented in *Table 10* (both taken from the submission). The primary OS analysis was performed with and without stratification. Stratification minimises the potential for bias by restricting comparisons to more homogeneous groups. Pre-specified stratification factors were: preselected CCR (IC versus LDAC or BSC); ECOG performance status (0–1 versus 2); and cytogenetic risk (intermediate versus poor).

Reporting mean OS may have offered superior understanding for the efficacy of treatments in comparison to what the median offers. It would have been better if Celgene had presented both the mean and median for OS.

Figure 2: Kaplan-Meier plot of overall survival



Key: CCR, conventional care regimens; CI, confidence interval; HR, hazard ratio.
Source: Celgene submission, Figure 8, p. 66

Table 10: Summary of overall survival in the ITT population

Outcome	Azacitidine (N=241)	CCR (N=247)
Event, n (%)	193 (80.1)	201 (81.4)
Censored, n (%)	48 (19.9)	46 (18.6)
Median OS (95% CI), months[a]	10.4 (8.0, 12.7)	6.5 (5.0, 8.6)
Difference (95% CI), months[a]	3.8 (1.0, 6.5)	
HR [AZA:CCR] (95% CI)[b]	0.85 (0.69, 1.03)	
Stratified log-rank test: p-value[c]	0.1009	
HR [AZA:CCR] (95% CI)[d]	0.84 (0.69, 1.02)	
Unstratified log-rank test: p-value[e]	0.0829	
1-year survival, % (95% CI)	46.5 (40.1, 52.7)	34.3 (28.3, 40.3)
Difference, % (95% CI)[f]	12.3 (3.5, 21.0)	

Key: AZA, azacitidine; CCR, conventional care regimens; CI, confidence interval; ECOG, Eastern Cooperative Oncology Group; HR, hazard ratio; ITT, intent-to-treat; KM, Kaplan Meier; OS, overall survival; PH, proportional hazards.

Notes: a, Median, 25th, and 75th percentile estimates of OS are from an unstratified KM analysis. Differences were calculated as AZA:CCR. The CIs for the differences were derived using Kosorok's method; b, The HR is from a Cox PH model stratified by ECOG performance status and cytogenetic risk status; c, p-value is two-sided from a log-rank test stratified by ECOG performance status, and cytogenetic risk status; d, the HR is from an unstratified Cox PH model; e, p-value is two-sided from an unstratified log-rank test; f, CI for the difference in the 1-year survival probabilities was derived using Greenwood's variance estimate.

Source: Celgene submission, Table 18, p. 66

4.2.2.4.2 Secondary efficacy endpoints

A summary of secondary endpoints for azacitidine versus CCR is presented in *Table 11*. The ERG used the data reported by Celgene to produce Kaplan-Meier figures for event-free survival (*Figure 3*), relapse-free survival (*Figure 4*) and progression-free survival (*Figure 5*). *Table 12* and *Table 13* present outcomes based on pre-selection prior to randomisation.

Table 11: Secondary endpoints: azacitidine versus CCR

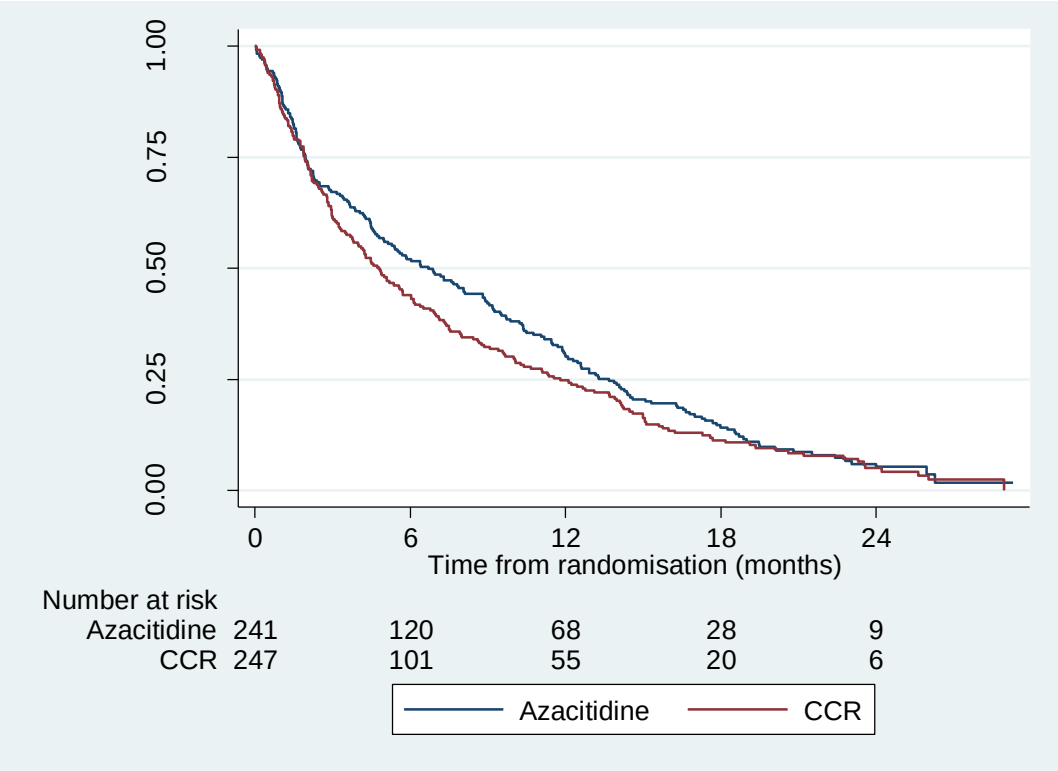
	Azacitidine (N=241)		CCR (N=247)		HR	95% CI	p value
	N	%	N	%			
Death estimates							
30-day	16	6.6	25	10.1	–	–	–
60-day	39	16.2	45	18.2	–	–	–
Haematologic response[a]							
CR + CRi	67	27.8	62	25.1	–	–	0.5384
CR	47	19.5	54	21.9	–	–	0.5766
CRc-20	5	2.1	14	5.7	–	–	0.0589
PR	3	1.2	3	1.2	–	–	1.0
Progressive disease	20	8.3	20	8.1	–	–	1.0
Stable disease	71	29.5	59	23.9	–	–	0.1833
Other secondary endpoints							
EFS[b]							
Median, months		6.7		4.8	0.87	0.72, 1.05	0.1495
RFS							
Median, months		9.3		10.5	1.11	0.75, 1.66	0.5832
Relapse after CR or CRi	43	63.2	35	56.5	–	–	0.4712
Duration of remission							
Median, months		10.4		12.3	–	–	–
Transfusion independence[c]							
RBC	65	38.5	45	27.6	–	–	–
Platelets	41	40.6	24	29.3	–	–	–

Key: CCR, conventional care regimens; CI, confidence interval; CR, complete response; CRc-20, complete cytogenetic remission in at least 20 metaphases; CRi, complete remission with incomplete blood count recovery; EFS, event-free survival; PR, partial remission; RBC, red blood cell; RFS, relapse-free survival.

Notes: a, Defined by International Working Group criteria and was adjusted by an independent review committee; b, Events included treatment failure, progressive disease, relapse after CR or CRi, or death; c, Defined as no transfusions for 56 consecutive days on study for patients who were transfusion dependent at baseline.

Source: Celgene submission, Table 22, p. 72

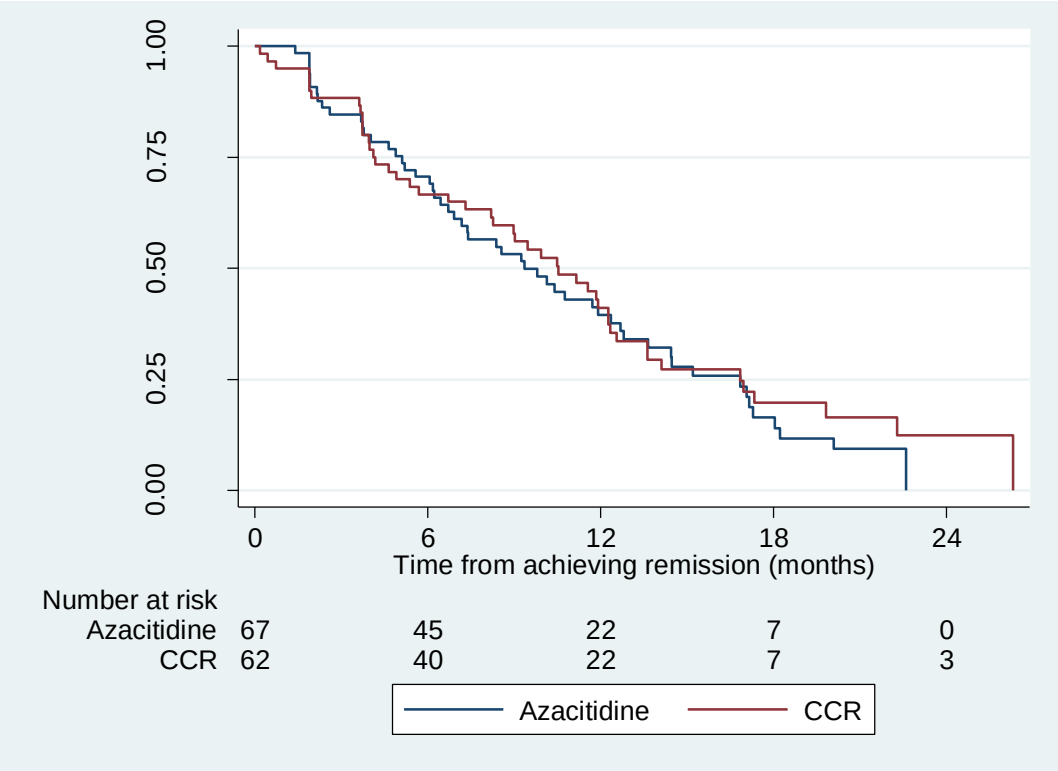
Figure 3: Kaplan-Meier plot of event-free survival



Key: CCR, conventional care regimens; CR, complete remission; CRi, complete remission with incomplete blood count recovery

Notes: Event-free survival, defined for all patients as the time from randomisation to treatment failure, disease progression, relapse after CR or CRi, death from any cause, or loss to follow-up

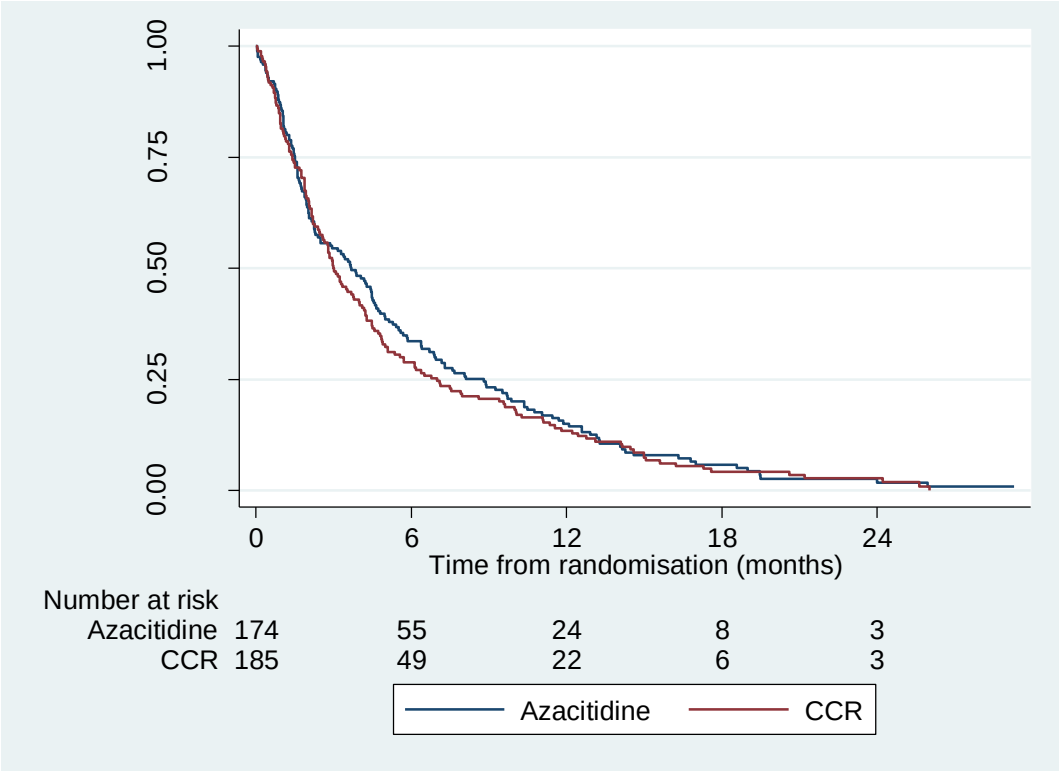
Figure 4: Kaplan-Meier plot of relapse-free survival



Key: CCR, conventional care regimens; CR, complete remission; CRi, complete remission with incomplete blood count recovery

Notes: Relapse-free survival, defined for patients achieving CR or CRi as the time from first documented CR or CRi to relapse, death from any cause, or loss to follow-up

Figure 5: Kaplan-Meier plot of progression-free survival



Key: CCR, conventional care regimens; CR, complete remission; CRi, complete remission with incomplete blood count recovery
Notes: Progression-free survival, defined as for event-free survival but for patients achieving neither CR nor CRi

Table 12: Secondary endpoints – according to investigator pre-selection: haematologic response

Investigator pre-selection	BSC		LDAC		IC	
	AZA (N=44)	CCR (N=45)	AZA (N=154)	CCR (N=158)	AZA (N=43)	CCR (N=44)
Haematologic response, n (%)[a]						
CR + CRi	7 (15.9)	0 (0.0)	42 (27.3)	41(25.9)	18 (41.9)	21 (47.7)
CR	6 (13.6)	0 (0.0)	28 (18.2)	38 (24.1)	13 (30.2)	16 (36.4)
CRc-20	1 (2.3)	0 (0.0)	3 (1.9)	8 (5.1)	1 (2.3)	6 (13.6)
PR	0 (0.0)	0 (0.0)	3 (1.9)	1 (0.6)	0 (0.0)	2 (4.5)
Progressive disease	4 (9.1)	5 (11.1)	10 (6.5)	14 (8.9)	6 (14.0)	1 (2.3)
Stable disease	14 (31.8)	6 (13.3)	47 (30.5)	46 (29.1)	10 (23.3)	7 (15.9)

Key: AZA, azacitidine; CCR, conventional care regimens; CR, complete response; CRc-20, complete cytogenetic remission in at least 20 metaphases; CRi, complete remission with incomplete blood count recovery; PR, partial remission.
Notes: a, Defined by International Working Group criteria and was adjusted by an independent review committee.
Source: Celgene submission, Table 23, p. 73; AZA-AML-001 Clinical Study Report[37]

Table 13: Secondary endpoints – according to investigator pre-selection: other secondary outcomes

Investigator pre-selection	BSC		LDAC		IC	
	AZA (N=44)	CCR (N=45)	AZA (N=154)	CCR (N=158)	AZA (N=44)	CCR (N=45)
Event-free survival						
Median, months	4.5	3.1	7.3	4.8	8.1	9.7
Hazard ratio (95% CI)	0.67 (0.43, 1.04)		0.89 (0.70, 1.13)		1.02 (0.64, 1.63)	
p-value	0.0756		0.3563		0.9196	
Relapse-free survival						
Median, months			8.6	9.9	10.8	12.1
Hazard ratio (95% CI)			1.11 (0.68, 1.81)		1.21 (0.58, 2.51)	
p-value			0.6638		0.6135	
Relapse after CR or CRi, n (%)	2 (28.6)	NA	31 (73.8)	25 (61.0)	10 (55.6)	10 (47.6)
Duration of remission						
Median, months			9.2	11.2	17.3	19.8

Key: AZA, azacitidine; CCR, conventional care regimens; CI, confidence interval; CR, complete response; CRi, complete remission with incomplete blood count recovery.
Sources: Celgene submission, Table 23, p. 73-74; AZA-AML-001 Clinical Study Report[37]

Similarly to the primary outcome, OS, reporting the mean for the secondary outcomes may have offered superior understanding for the efficacy of treatments in comparison to what the median offers. There is also limited reporting of time-to-event reporting for the secondary outcomes.

One-year survival

Comments on one year survival from Celgene were as follows (Source: Celgene submission, Section 4.7.5.1, p. 75):

> *Azacitidine improved 1-year survival compared with CCR (46.5% vs. 34.3%, respectively), resulting in a clinically meaningful difference of 12.3% in favour of azacitidine (95% CI: 3.5, 21.0).*
>
> *1-year survival was also improved for azacitidine when compared with each of the CCR therapies (within investigator pre-selection) (BSC only: 30.3% vs. 18.6%, LDAC: 48.5% vs. 34.0%, and IC: 55.8% vs. 50.9%, respectively), and in a post-hoc analysis, when compared with BSC plus LDAC (within investigator pre-selection) (44.5% vs. 30.6%, respectively).*

None of these improvements reported by Celgene were significantly different.

Event-free survival

Additional information about event-free survival from Celgene was given (Source: Celgene submission, Section 4.7.5.3, p. 75):

> *Overall, 212 (88.0%) events (defined as treatment failure, progressive disease, relapse after CR or CRi, death from any cause, or loss to follow-up) were reported in subjects treated with azacitidine and 216 (87.4%) events in subjects treated with CCR.*

4.2.2.4.3 Health-related quality of life

The European organisation for research treatment of cancer (EORTC) QLQ-C-30 was used to assess HRQoL. The questionnaire was completed at baseline, on day 1 of every other cycle and at the end-of-study visit. To be evaluable as the HRQoL population, baseline assessment and at least one other post-baseline assessment was required. The HRQoL population comprised of initially of a total of n=291 (n=157 azacitidine and n=134 CCR). The size of the HRQoL population decreased in size over time in both groups. *Table 14* presents the evaluable HRQoL population throughout treatment up to cycle 9.

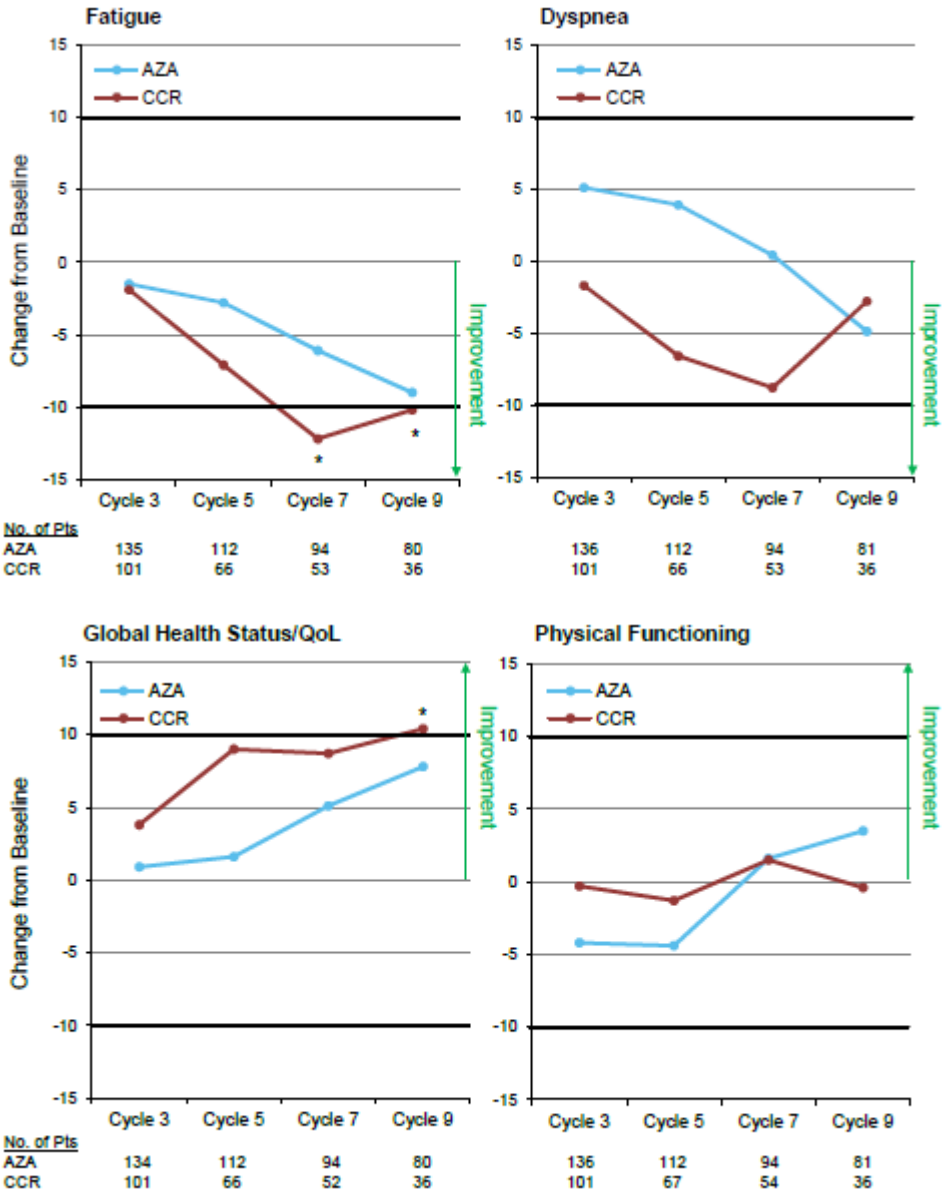
Primary HRQoL endpoints reported include fatigue score, dyspnoea, physical functioning and global health status. Changes over time for these four endpoints are depicted in *Figure 6*.

Table 14: HRQoL assessment rates

HRQoL assessment	AZA (n=241)			CCR (n=247)		
	Treated	Assessed n (%)	Evaluable n (%)	Treated n	Assessed n (%)	Evaluable n (%)
Cycle 1 (Baseline)	237	210 (89)	157 (66)	236	210 (89)	134 (57)
Cycle 3	174	152 (87)	137 (79)	131	113 (86)	102 (78)
Cycle 5	146	127 (87)	112 (77)	86	72 (84)	67 (78)
Cycle 7	118	105 (89)	94 (80)	67	58 (87)	54 (81)
Cycle 9	98	89 (91)	81 (83)	49	38 (78)	36 (73)

Key: AZA, azacitidine; CCR, conventional care regimens; CI, confidence interval; HR, hazard ratio.
Notes: HRQOL assessment rates = number of patients with an EORTC QLQ-C30 assessment ÷ the total number of patients receiving treatment at the scheduled cycle visit. Numbers reported in this table represent all HRQOL assessments at each cycle; some patients may not be included in HRQL analyses due to missing baseline HRQOL assessments. Evaluable patients completed an HRQL assessment at baseline and had ≥ 1 post-baseline assessment.
Source: Celgene submission, Table 24, p. 77

Figure 6: Mean absolute score change from baseline for primary and secondary HRQoL endpoints (HRQoL evaluable population)



Key: AZA, azacitidine; CCR, conventional care regimens.
Notes: Decreasing scores indicate improvement in the Fatigue and Dyspnoea domains of the QLQ-C30, and increasing scores indicate improvement in the Physical Function and Global Health Status/QoL domains.
The minimally important difference, defined as a mean change of at least 10 points from baseline and representing a clinically meaningful effect is denoted by bold black lines at 10 and -10 on the y-axis.
*Met the threshold for minimally important difference.
Source: Celgene Submission, Figure 10, p. 78

Statistical tests were not reported between treatment arms for HRQoL. From Figure 6, the CCR arm appears to be favourable in comparison to azacitidine for cycle three, five and seven the four HRQoL measures reported (fatigue, dyspnoea, global health status and physical functioning). For the final cycle reported (cycle nine), CCR is favourable for fatigue and global health status, whilst azacitidine is favourable for physical functioning and dyspnoea.

There are two significant critiques of the HRQoL analyses by Celgene:

- Assessments are made at the start of each cycle, after a significant recovery period for patients after treatment;
- A significant number of patients (197 patients, 40%) are not represented in the HRQoL assessment at any time point as they did not have a baseline assessment (14%) and/or lacked a post-baseline assessment; these patients are likely to have been more ill (lower HRQoL) than the patients who were assessed and evaluable.

4.2.2.4.4 Adverse events

AML-AZA-001 included 471 subjects in the safety population who had newly diagnosed AML with >30% blasts and were randomised to receive azacitidine or CCR. The median age of the safety population was 75.0 years with 53.3% of subjects ≥75 years of age, 32.7% of subjects had AML with MDS-related changes, 18.0% had a prior history of MDS, 35.7% had a poor or very poor cytogenetic risk status, and 22.9% had ECOG performance status of 2. Median baseline BM blast count was 71.5%.

A summary for adverse events (AEs) is presented in *Table 15*. *Table 16* reports the incidences of AEs for > 10 % of people in any treatment arm. *Tables A3*, *A4* and *A5* (*Appendix 1*) presents AEs for both treatment arms, based on the number of events occurring in ≥10% of people in the azacitidine group.

Table 15: Summary of adverse events

Adverse events	AZA, n (%) (N=236)	CCR		
		BSC, n (%) (N=40)	LDAC, n (%) (N=153)	IC, n (%) (N=42)
≥1 AE	234 (99.2)	36 (90.0)	153 (100.0)	42 (100.0)
≥1 treatment-related AE	188 (79.7)	0 (0.0)	124 (81.0)	39 (92.9)
≥1 Grade 3 or 4 AE	207 (87.7)	26 (65.0)	141 (92.2)	37 (88.1)
≥1 Grade 3 or 4 treatment-related AE	125 (53.0)	0 (0.0)	90 (58.8)	29 (69.0)
≥1 Grade 5 (leading to death) AE	56 (23.7)	23 (57.5)	38 (24.8)	9 (21.4)
≥1 Grade 5 (leading to death) treatment-related AE	12 (5.1)	0 (0.0)	10 (6.5)	4 (9.5)
≥1 SAE	188 (79.7)	30 (75.0)	118 (77.1)	27 (64.3)
≥1 treatment-related SAE	87 (36.9)	0 (0.0)	56 (36.6)	14 (33.3)
≥1 AE leading to discontinuation	110 (46.6)	0 (0.0)	68 (44.4)	11 (26.2)
≥1 treatment-related AE leading to discontinuation	22 (9.3)	0 (0.0)	20 (13.1)	5 (11.9)
≥1 AE leading to dose reduction only	8 (3.4)	0 (0.0)	2 (1.3)	2 (4.8)
≥1 AE leading to study drug dose interruption only	116 (49.2)	0 (0.0)	61 (39.9)	4 (9.5)
≥1 AE leading to study drug dose reduction and interruption	13 (5.5)	0 (0.0)	7 (4.6)	0 (0.0)

Key: AE, adverse event; AML, acute myeloid leukaemia; AZA, azacitidine; BSC, best supportive care; CCR, conventional care regimens; CI, confidence interval; IC, intensive chemotherapy; LDAC, low-dose cytarabine; SAE, serious adverse event.

Notes: AE refers to treatment-emergent adverse events. Adverse events included events that started (1) between the date of first dose of study drug and 28 days after the date of last dose of study drug for azacitidine and LDAC (2) between the date of first dose of study drug and 70 days after the date of last dose of study drug for IC (3) between the date of randomisation and the date of discontinuation from the treatment period for BSC only. Adverse events that started outside the treatment-emergent period and assessed as related to study drug was considered treatment-emergent.

Source: Celgene Submission, Table 26, p. 86

Table 16: Grade 3 to 4 Treatment emergent adverse events occurring in ≥10% of patients in any treatment group

Adverse events	Azacitidine (n=236)		CCR					
			BSC only (n=40)		LDAC (n=153)		IC (n=42)	
Preferred term	No.	%	No.	%	No.	%	No.	%
Febrile neutropenia	66	28.0	11	27.5	46	30.1	13	31.0
Neutropenia	62	26.3	2	5.0	38	24.8	14	33.3
Thrombocytopenia	56	23.7	2	5.0	42	27.5	9	21.4
Pneumonia	45	19.1	2	5.0	29	19.0	2	4.8
Anaemia	37	15.7	2	5.0	35	22.9	6	14.3
Leukopenia	16	6.8	0	0	13	8.5	6	14.3
Hypokalaemia	12	5.1	1	2.5	10	6.5	7	16.7

Key: BSC, best supportive care; CCR, conventional care regimens; IC, intensive chemotherapy; LDAC, low-dose cytarabine.
Source: Dombret et al. 2015[45]

4.2.3 Interpretation

Key efficacy findings from the RCT reported from the submission were as follows:

Overall survival

Azacitidine was not significantly superior to CCR in prolonging survival of adults ≥65 years with AML with >30% bone marrow blasts.

Secondary endpoints

1-year survival rates were 46.5% for azacitidine compared to 34.3% in the CCR arm (difference 12.3 %; 95% CI: 3.5, 21.0).

Measures of haematologic response, duration of remission and remission free survival were similar between treatment arms when CCR was combined. When CCR was not combined, it appeared (although limited statistical analysis was reported) that IC was superior to azacitidine. Participant numbers for the IC arm compared to those originally assigned to IC were small (n=44 and n=43 respectively).

No statistical analyses were presented for the HRQoL data. Appearances from the figures suggest that CCR was favourable to azacitidine.

Adverse events

Treatment related AEs were common for both azacitidine, LDAC and IC. Unsurprisingly, AEs were less common for BSC.

4.2.3.1 Strengths and limitations

Strengths

- Multicentre, appropriately randomised design of the RCT AZA-AML-001

- The population recruited to AZA-AML-001 was representative of the typical UK patient population
- The appropriate CCR regimes used as a comparator in AZA-AML-001 compared to UK standard practices

Limitations

- Underpowered for individual CCR arms
- The open-label design introduces the risk of bias
- The use of subsequent therapies following treatment assignment, some of which are not used in routine NHS practice, and which were not balanced across treatment arms
- Limited reporting of time-to-treat outcomes (except for OS)
- Use of statistical analyses which have reduced efficiency when proportional hazard assumptions are not met

4.3 Adjustments of overall survival estimates for subsequent therapy

In order to address confounding effects of subsequent therapy on overall survival (OS), the company's submission presents post-hoc analyses that adjust for subsequent therapy use. Among these, the analyses that censored data at the start of subsequent therapy and weighted the remaining data by the inverse of the probability of not being censored, i.e., the inverse probability of censoring weights (IPCW) method, play the most prominent role in the submission. Other methods of adjustment of OS treatment effects were considered elsewhere in the submission (Celgene submission, Section 5.3.5, p. 122), such as the Rank Preserving Structural Failure Time Model (RPSFTM) and the Iterative Parameter Estimation (IPE), but these were considered inferior to the IPCW.

The company presented estimates of relative effects for two sets of IPCW analyses. In the first set of analyses, adjustment for subsequent treatment was applied to both trial arms (i.e., azacitidine and CCR). In the second set of analyses, adjustment for subsequent treatment with azacitidine was applied to the CCR arm.

Celgene adopted the results of the second set for its health economic base case analysis. The company justified this choice on the basis of methodological guidelines (Source: Celgene submission, Section 4.4.4.3, p. 55):

> *For regulatory purposes, an initial IPCW analysis was undertaken in which both treatment arms were adjusted. A further IPCW analysis was conducted in line with the NICE DSU TSD16 in which adjustments were only made to the comparator treatment arm (CCR).*[51]

This reasoning led Celgene to devote little space in its submission to describe the first set of IPCW analyses, which adjusted for subsequent therapy in both arms, and to focus on detailing the second IPCW analysis, which only adjusted for subsequent azacitidine use in the CCR arm.

The ERG believes the first set of analyses, which adjust for subsequent AML treatment in both trial arms, should have been used for the base case economic analysis instead of the IPCW analysis adjusting for azacitidine use in CCR. The ERG believes the company are mistaken in their interpretation of the methodological guidelines. The NICE DSU TSD16 illustrates the application of IPCW to estimate treatment effects in a hypothetical context where a policy choice needs to be made among two states of the world, one where a new, experimental treatment is available to patients and the alternative state where such treatment is not available, and the evidence comes from an RCT where some control patients switch to the experimental treatment after randomisation. IPCW adjustment (or indeed adjustment by any other suitable method for that matter) of outcome data in the control arm would then be warranted. In this example no IPCW adjustment was made in the experimental arm because it was assumed that the mix of subsequent therapies used by patients in that arm is representative of patterns of care in the state of the world where the experimental treatment is made available. In contrast, as acknowledged in the NICE DSU TSD16, if the mix of subsequent therapies in any trial arm (whether is the experimental or control) is not representative of patterns of care in the state of the world of interest to the decision problem, then adjustment for those subsequent treatments is needed.

Therefore, in the case where, as the cost-effectiveness model submitted by Celgene for this appraisal implies (see *Section 5*), the decision problem involves the evaluation of two alternative states of the world where no subsequent active treatment is available, then adjustment of outcomes for active treatment switching in both trial arms is warranted. Whether the premise that no subsequent active treatment would be available to UK patients in routine practice is correct may of course be questioned (see *Section 5* for clinical expert opinion on this issue obtained by the ERG for this appraisal), but the point here is that applying IPCW adjustment for subsequent treatment to the outcome data of the CCR but not the azacitidine arm in the AZA-AML-001 trial is inconsistent with the economic model which the OS IPCW analysis was designed to inform.

Since OS differences between the two arms in AZA-AML-001 narrowed over time and the statistically insignificant ITT findings in OS were reversed when data were censored at subsequent AML treatment initiation (Celgene submission, Section 4.7.2.3, pp. 67–68), Celgene undertook further post-hoc analyses to adjust for imbalance in the use of subsequent treatment and baseline covariates across trial arms. The following methods were explored.

4.3.1 Cox proportional hazards models

Three post-hoc Cox proportional hazards models of survival were considered in the clinical effectiveness submission, although were not part of scenarios investigated in the economic submission:

- Model 1 was a function of a time varying indicator of subsequent treatment (as a main effect and interacted with treatment allocation);
- Model 2 was a function of baseline covariates alone;
- Model 3 included covariates in Models 1 and 2.

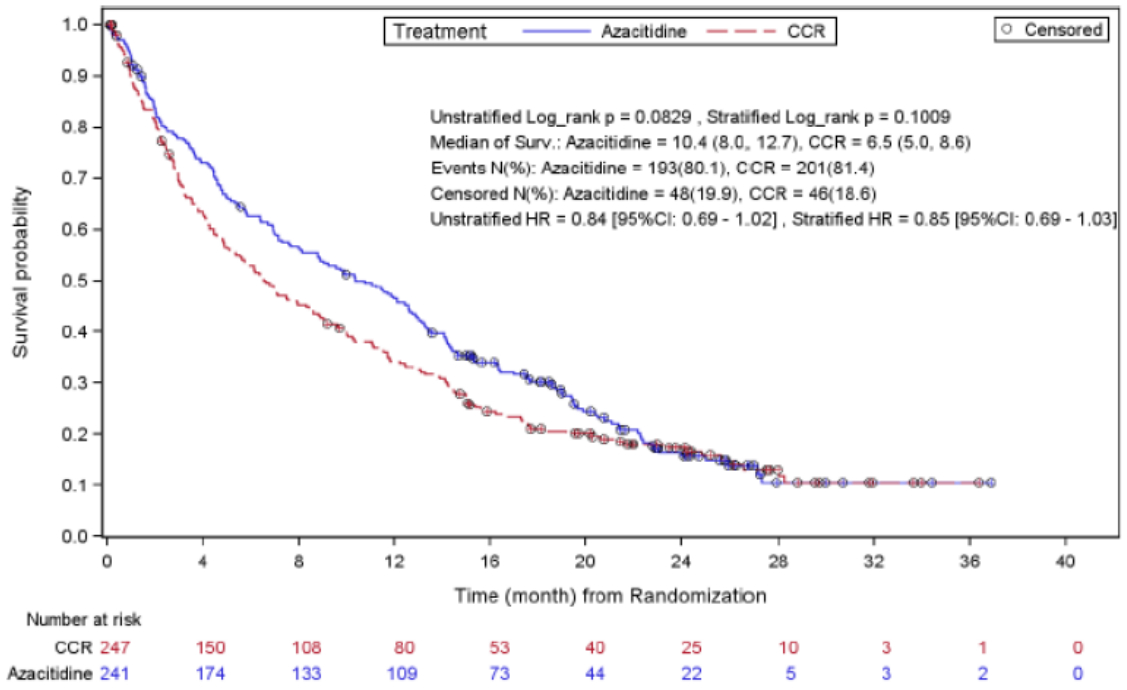
Model 3 produced a large treatment effect estimate (HR 0.69, 95% CI: 0.54, 0.88; p=0.0027). However, this estimate and those from the other two post-hoc Cox proportional hazards models are susceptible to bias.

The treatment effect of azacitidine in Model 2 is likely confounded by treatment switching, whilst the adjustment for treatment switching in Model 1 and Model 3 implausibly implies the following two assumptions: (a) either those who switch have the same prognosis as those that do not switch or their prognoses differ but they are evenly distributed across arms; and (b) subsequent treatments have the same average effect across arms conditional on prognosis. The different mix of subsequent therapies used across azacitidine and CCR arms (*see Table 48, page 105*) and noticeable differences in results between Model 3 and the respective IPCW analysis (*see Section 4.3.2*) suggests that neither assumption (a) nor (b) are likely to be borne out by the data.

4.3.2 Inverse probability of censoring weights (IPCW) method

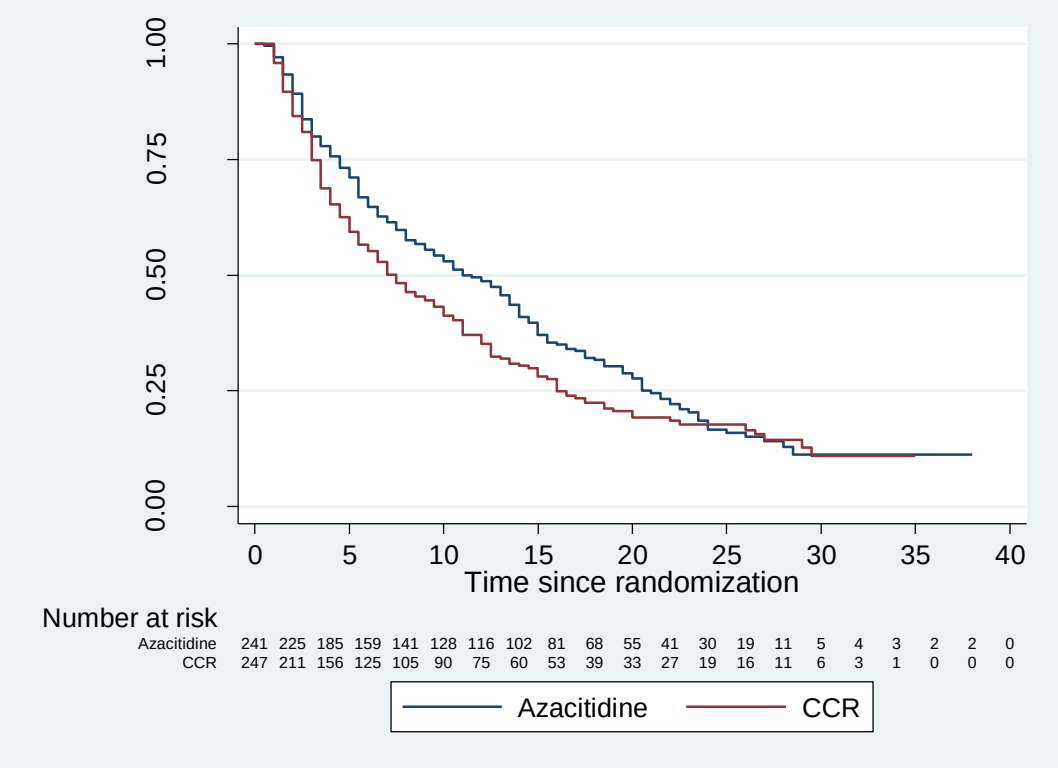
The IPCW method was used, adjusting for treatment switching in the CCR arm only. This approach sought to account for the possibility that subsequent treatment use did not occur at random. Results were presented for a Cox proportional hazards model unadjusted for differences in baseline characteristics across treatment arms (unadjusted IPCW Cox proportional hazards model; HR AZA versus CCR [redacted], 95% CI: [redacted], and another Cox proportional hazards model that included covariates for those baseline characteristics (adjusted IPCW Cox proportional hazards model; [redacted], [redacted]. The adjusted IPCW Cox proportional hazards estimate was used in the base case economic analysis by Celgene (*see Section 5.2.5, page 73*). This estimate only adjusts outcomes in the CCR arm for subsequent azacitidine use, which accounted for 32 out of the 74 (43.2%) control subjects with subsequent treatment (Source: Celgene submission, Section 4.7.2.3, p. 67). *Figure 7* and *Figure 8* show the Kaplan-Meier curves for the ITT and IPCW data, measured in days and 15-day periods respectively (Celgene did not provide the figure in days or the individual patient data to produce this graph to ERG). The azacitidine curve is the same in the two figures, whereas the CCR curve with IPCW lies above that of ITT.

Figure 7: Overall survival in AZA-AML-001 based on intention-to-treat population



Key: CCR, conventional care regimen; HR, hazard ratio
Source: Celgene submission, Figure 8, p. 66

Figure 8: Overall survival in AZA-AML-001 following adjustment of the CCR arm for subsequent treatment with azacitidine using the IPCW method



Key: CCR, conventional care regimen; IPCW, inverse probability of censoring weight
Note: Based on discrete (15-day period) time-to-event data
Source: Patient-level data supplied to ERG by Celgene

The company also reported that "For regulatory purposes, an initial IPCW analysis was undertaken in which both treatment arms were adjusted." (Source: Celgene Submission, Section 4.4.4.3., p. 55). Two sets of results for these IPCW Cox proportional hazards models of treatment switching in both AZA and CCR treatment arms were presented, one set for a model that adjusted for baseline prognostic covariates and another for a model unadjusted for those covariates (Source: Celgene submission, Table 21, p. 70). The baseline covariate-unadjusted IPCW Cox proportional hazards model HR estimate was 0.77 (95% CI, 0.61–0.98), and the adjusted estimate was 0.71 (95% CI, 0.56–0.90) (Celgene submission: Table 21, p. 70).

Although the respective methods are not clearly reported, it appears that these estimates adjust for any subsequent treatment in both arms, as opposed to adjusting only for azacitidine in the CCR arm as in the IPCW Cox proportional hazards method previously described and used by Celgene for its base case economic analysis. As explained in *Section 5,* if IPCW methods were indeed used to adjust for any subsequent treatment in both arms, the associated results may be the most suitable of those submitted by Celgene to populate the model with for the company's base case economic analysis. However, the submission is unclear about what treatments these analyses adjusted for (Source: Celgene, Appendices to submission, Appendix 11, Section 3.3.1):

[redacted]

[redacted]

In AZA-AML-001, azacitidine was not the only subsequent treatment for AML used in the CCR arm, and there were also active treatments after azacitidine in the azacitidine arm (*see Table 48, page 105*).

4.3.3 Other analyses

Other analyses submitted by the company separately analysed patients who did and did not receive any subsequent therapy. These analyses are by their nature of limited use in informing assessments of relative effectiveness, because the differences in outcomes between those two groups of patients are likely to confound the effect of treatment with individual variation in the propensity to receive treatment, and are not reviewed further.

Another set of treatment effect estimates were reported for an analysis that censored subsequent cytarabine-based therapy in the azacitidine group and subsequent azacitidine in the CCR group (HR [redacted]; 95% CI: [redacted]). These results are more in line with the assumptions of Celgene's economic model, which assumes no subsequent active treatment after treatment failure is used in the azacitidine and CCR groups. However, as reported by the Celgene submission, cytarabine accounted only for 40 (59.7%) of the 67 subjects who received subsequent treatment in the azacitidine group (Celgene submission, Section 4.7.2.3, p. 67).

Another analysis used an imputation method to adjust for treatment switching (Celgene Post hoc statistical methods addendum AZA-AML-001), but since little information was provided on the methods and the results of this analysis played no subsequent role in the submission, we do not discuss these further.

4.3.4 Summary and additional issues

Comparing HR estimates between the baseline covariate-unadjusted IPCW PH, 0.77 (95% CI: 0.61–0.98) (Celgene submission, Table 21, p. 70), and the unstratified HR (ITT) censoring at switch to AML therapy, 0.75 (95% CI: 0.59–0.95) (Celgene submission, Table 19, p. 68), suggests sophisticated methods of adjustment for subsequent treatment use that account for censoring not-at-random (e.g., IPCW) make little difference versus simple methods that assume censoring at random.

Further, similar HR estimates were found with IPCW Cox proportional hazards versus simply estimating Cox proportional hazards treatment effects without subsequent therapy-related censoring and a time-varying covariate indicator of subsequent therapy interacted with randomly allocated therapy (0.71 versus 0.69).

Adjusting versus not adjusting for baseline covariates appears to be the single structural factor to which estimates of relative effectiveness are most sensitive, as evidenced by the difference in HR estimates between Cox proportional hazards Model 1 versus Model 3 (0.75 versus 0.69) and the corresponding estimates with IPCW applied to both trial arms (0.77 versus 0.71).

The wide range of HR estimates reported by the company relies on the assumption of constant proportional hazards, which was statistically tested for the Cox but not for the IPCW approaches. No reasons for the absence of tests of the proportional hazards assumption in the IPCW analyses were presented in the submission. The methods description presented in the cost-effectiveness section of the Celgene submission suggests these tests were not performed and does not provide reasons for this omission (*see Section 5.3* for our critique of methods presented in the cost-effectiveness section of the Celgene submission).

In *Figure A1* (*Appendix 2*) we present log-log plots for the IPCW analysis adjusting for azacitidine use in the CCR arm only (the corresponding data for the IPCW analysis adjusting both the azacitidine and CCR arm for subsequent AML therapy use were not provided by the company). This suggests that the proportional hazards assumption (i.e., curves being parallel) is unlikely to hold after month 20. Statistical test of this assumption using Schoenfeld residuals also rejects the assumption (X^2=5.82; p=0.016).

As explained in *Section 5.3.6.2* (*page 111*), the IPCW method applied by Celgene consisted of two parts, first a regression analysis to predict the probability of a CCR patient not receiving subsequent treatment at each follow-up point (every 15-days), as a function of fixed baseline covariates and time-varying covariates that affect both the likelihood of treatment switching and overall survival outcomes. In a second step the survival time of CCR individuals who switched to azacitidine are censored at the time of switching and the inverse of the predicted probability of switching corresponding to uncensored CCR individuals at a given follow-up point is used as sampling weight to account for the unobserved outcomes of censored individuals in the counterfactual situation that they had not switched treatment. This method thus intends to estimate the outcomes that would have been observed in the CCR sample had no patients switched to azacitidine, and does so by adjusting for non-random censoring (i.e., treatment switching) using a predictive model for subsequent treatment.

The validity of overall survival IPCW estimates (and resulting hazard ratios) depends on the AZA-AML-001 data conforming to two key assumptions.

The first key assumption (referred to as the exchangeability assumption) dictates that after controlling for the measured predictors common to survival and subsequent azacitidine use, subsequent azacitidine users have the same survival prognosis as those who remain in CCR. As discussed elsewhere, this assumption only holds under the following three conditions:

- All common predictors are appropriately measured and accounted for in the analysis;
- The sample of patients available at all follow-up times is sufficiently large to ensure that the probability of not switching is positive for every combination of values observed for the common predictors over the whole sequence of follow-up points (i.e., the positivity condition, Cole and Hernan 2008[52]);

- The common predictors cannot perfectly or nearly perfectly predict survival or switching to azacitidine.

Thus, small sample size or perfect common predictors violate the exchangeability assumption because the survival outcomes of uncensored CCR patients would not be representative of censored individuals, who switched to azacitidine, even if all common confounders were appropriately measured and included in the analysis.[53] This problem is thus aggravated by small sample sizes, highly stratified data due to a large set of predictors used to estimate weights, and continuous predictors can generate random non-positivity.

The second assumption is that the functional forms in which the common predictors enter the switching prediction model is correctly specified, so that exchangeability is maximised by controlling for selection bias while maintaining positivity.[52]

While the ERG could replicate the IPCW hazard ratio estimates reported by Celgene using individual patient data provided by the company, it was not provided the data to replicate the estimation of IPCW weights, which were estimated in SAS (Celgene appendices to submission, Appendix 10). This limited the ability to assess whether the assumptions implied by the IPCW method are consistent with the data from AZA-AML-001. Further details are provided in *Section 5.3.6.2* (*page 111*).

Celgene notes that besides the limitations inherent in the IPCW, there was an additional limitation (Source: Celgene Post hoc statistical methods addendum AZA-AML-001, pp. 6–7):

> *An additional limitation of the post-randomization data in the current study was that for subjects who did not receive subsequent therapy, the last visit interval extended from the last assessment done at the time of discontinuation from the treatment phase of the study to the time of censoring (study closure or lost to follow-up) or death. If the subject was alive and in the survival follow-up phase for a long period of time then the gap between treatment discontinuation and death or censoring could be quite long and represents a time interval during which no additional clinical information was collected or available.*

This issue is a potentially important one for the company's IPCW analyses, because it suggests that an unknown proportion of patients might have received unrecorded subsequent therapy, for which survival outcomes remained unadjusted for. For the IPCW analysis that adjusted only outcomes of CCR arm, this implies that the benefit of azacitidine may be underestimated. In contrast, in the IPCW analysis that adjusted survival outcomes of both trial arms, if patients with improved survival prognosis were to be more likely to have received subsequent therapy during the period of unrecorded clinical management activity it is possible that unrecorded treatment switching may have produced a larger omitted variable bias in overall survival outcomes of azacitidine than CCR, and that the survival benefits of azacitidine may be overestimated.

The implications of these issues are covered further in *Section 5* that discusses the cost-effectiveness analysis submitted by Celgene.

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

As one RCT was identified from the searches and screening,[45] indirect comparisons and/or multiple treatment comparisons could not be made and were not presented by the company.

4.5 Critique of other evidence sources

Celgene provided four further sources (one non-RCT and three trial registries) of evidence within their submission.

4.5.1 Non-randomised evidence

The single non-RCT relevant to the decision problem identified by Celgene was by Lao et al.[46] This was a study which included a broader population of >20% blasts, however they provided sub-analysis for the population with >30% blasts. Celgene do not report any of the findings from this study as they claim the number of people with >30% blasts was small (n=12 for azacitidine group versus n=8 for the IC group versus n=22 for BSC group). From exploring the Lao et al. study, there is no evidence reported that compares any outcomes between the azacitidine group and either IC or BSC for those with >30% blasts. There is evidence comparing <30% and 30% or higher BM blasts in the azacitidine arm.

4.5.2 Registry data

The company report registry data from three countries: Austria, Spain and France. They hoped to conduct matched-adjusted indirect comparison using single-arm data. None of the published registry data (Austrian azacitidine registry (NCT01595295); Spanish AMLA registry and French compassionate patient named programme) provided information on the population with >30% bone marrow blasts, a population requirements for the Scope for this report.

4.6 Conclusions of the clinical effectiveness section

In spite of poorly designed and reported searches, the company submission identified a single RCT trial (AZA-AML-001) that matched the decision problem. Information from this study was reported in detail. The RCT was well conducted, although underpowered for each of the CCR arms.

The primary efficacy endpoint for the RCT was an ITT comparison of overall survival for patients randomised to azacitidine versus patients randomised to CCR. An improvement was demonstrated but it did not reach statistical significance. Statistical significance was also not reached for other outcomes assessed and reported. There was some evidence, though not statistically significant due to underpowering, that azacitidine was inferior to LDAC and IC on a number of outcomes, including response rate and relapse-free survival, although azacitidine appeared to be superior in relation to overall survival.

Post-hoc analyses were presented where overall survival outcomes were adjusted for subsequent treatment in CCR, where a significant improvement with azacitidine was found. However, these analyses were hampered by the fact that they relied on the assumption that treatment effects displayed a proportional hazards pattern, which was not statistically tested in the analysis sample.

5 Cost-effectiveness

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

5.1.1 Objective

The company conducted a systematic review to identify cost-effectiveness studies from the published literature relevant to the decision problem (*see Section 3, page 23*).

5.1.2 Search strategies

5.1.2.1 Economic evaluations

The company presented a literature search protocol to support its review of cost effectiveness. This protocol included systematic searches of key biomedical databases using a literature search strategy combined with hand-searching of conference abstracts and included studies. The literature searching was last updated in October 2015.

The bibliographic database searching used a search strategy that took the following form:

1. (search terms for acute myeloid leukaemia/leukaemia); *and*
2. (search terms for Azacitidine, cytarabine, clofarabine, daunorubicin, etoposide, fludarabine, idarubicin, mitoxantrone, mercaptopurine, amsacrine, cytotoxic or anthracycline or supportive or conventional care or chemotherapy or antineoplastic agent); *and*
3. (search terms to identify cost analysis, studies reporting economic evaluations or cost parameters).

This search strategy was applied in: MEDLINE and EMBASE (both via OVID), NHS EEDs and the HTA library (via the Cochrane Library: OVID interface) and Econlit (OVID).

The following conference proceedings were hand-searched between 2013 and 2015 inclusive:

- International Society for Pharmacoeconomics and Outcomes Research (ISPOR) - European, International, Asia-Pac, and Latin American
- American Society of Clinical Oncology (ASCO)
- European Society for Medical Oncology (ESMO)
- Tufts Cost-effectiveness Analysis (CEA) registry
- WHO International Clinical Trials Registry Platform (ICTRP)
- HTA Database of the International Network of Agencies for Health Technology Assessment (INAHTA)
- NIHR HTA website
- Research Papers in Economics (RePEc) website
- NICE website.

The ERG accepts that these literature searches were fit for purpose.

5.1.2.2 Health-related quality of life

Celgene did not record a systematic search for studies reporting HRQoL in their original submission. In response to a question for clarification, Celgene confirmed that they had undertaken a targeted systematic search. This literature search was undertaken in PubMed and the HERC database and takes population search terms combined with search terms for: QALY or utlilit* (search term is truncated) or EQ-5D. The ERG has been able to replicate and validate this search.

Celgene noted further that HRQoL was recorded for all treatments in the trial that supports their submission. Celgene have therefore used this data to parameterise their model arguing that it best reflects the quality of life seen in the target population. Given the inherent difficulty of measuring HRQoL outcomes in oncology trials and in particular in this clinical area, ERG believes that Celgene is likely to have identified the best available source of evidence on these outcomes relevant to this assessment.

5.1.3 Inclusion criteria

The review included cost-utility studies in the English language, or English abstracts of studies from studies written in other languages, relating to adults (>18 years) with AML with >30% blasts. In terms of treatments, included studies had to investigate azacitidine (intervention) and comparators of including intensive or non-intensive low-dose chemotherapy, including:

- Cytarabine (Ara C)
- Clofarabine (Evoltra®)
- Daunorubicin (daunomycin)
- Etoposide (Etopophos®, Vepesid®)
- Fludarabine (Fludara®)
- Idarubucin (Zavedos®)
- Mercaptopurine (Xaluprine®)
- Mitoxantrone
- Amsacrine
- Hydroxycarbamide (Hydrea)

or best supportive care (BSC).

All of these criteria were within the NICE Scope.

5.1.4 Results

Of 334 titles identified by the company's searches, forty eight studies were candidates for inclusion based on abstract and title content and its full text screened for possible inclusion. Forty studies were excluded after screening their full text, mostly due to the patient population investigated; six studies were excluded due to the study design being limited to

costs only rather than cost-effectiveness analysis. The remaining eight studies were included in the review.

Of the eight reviewed studies four were published abstracts and the remaining four were full journal article publications. Two of the full publications reported studies in populations within the NICE score of AML with >30% blasts, one was a study conducted in France,[54] and another was a study conducted in China.[55] The company performed quality assessment of the four full publications only, since abstract publication reported insufficient information to be assessed.

Although Celgene did not discuss the results of their quality assessment of the four included full publications, the quality checklist (Celgene appendices to submission, Appendix 8) makes clear that the studies were of very poor quality in terms of design, since no study was designed as an incremental cost-utility analysis, details of modelling methods required to extrapolate outcomes from short term efficacy studies to capture all important costs and benefits were not provided, and the degree of uncertainty in estimates was not measured. None of the studies were conducted in the UK and the one European study identified did not provide adequate methodological evidence to be of use to this assessment.

Only one of the reviewed studies evaluated azacitidine (versus LDAC). This was a study conducted in Russia that reported that azacitidine resulted in RUB 909,573 ($32,261 at 2012 PPP; source: OECD http://www.oecd-ilibrary.org/economics/purchasing-power-parities-for-gdp-2013-1_ppp-gdp-table-2013-1-en) extra costs and 0.76 additional life years per patient than LDAC, for an incremental cost per life year gained of RUB 1,196,808 (£42,449), but since it was published as an abstract[56] it provided limited information (ICER calculations made by ERG from information reported in the publication). Further, its relevance to the NICE decision problem is ambiguous since it is not clear whether patients had >30% blasts, and it included MDS patients. Moreover, these results did not take account of HRQoL outcomes and the costs results may not be transferable due to differences in relative prices between UK and Russia.

5.1.5 Conclusions

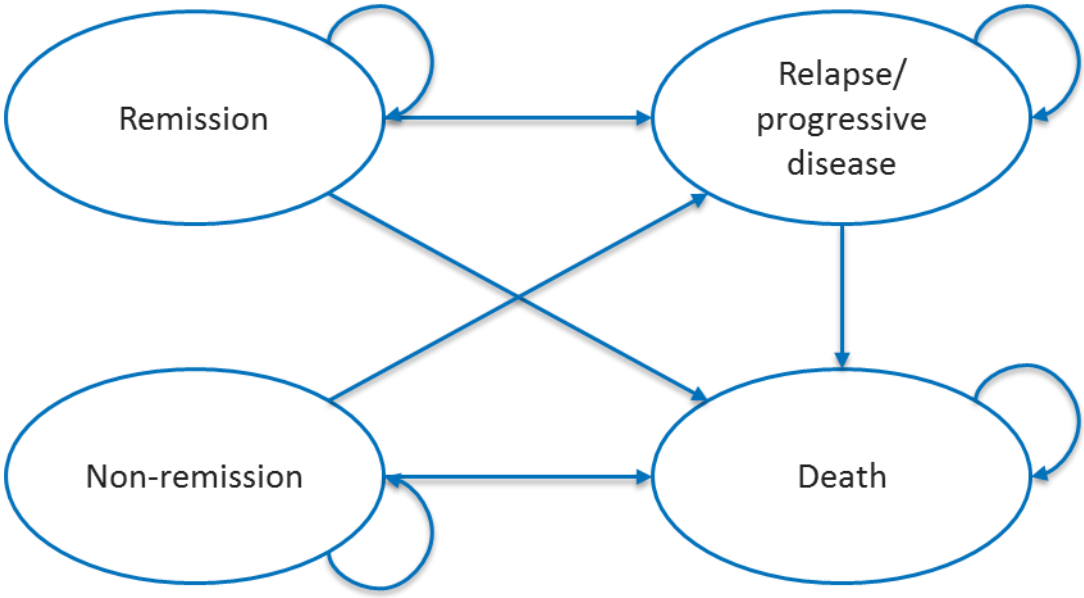
Although no conclusions were provided by the submitted review, ERG concludes that the quality of the evidence is poor and in any case unlikely to be relevant to the present assessment.

5.2 Summary of company's submitted economic evaluation

5.2.1 Model structure

A semi-Markov model, proposed by Celgene, is based on AZA-AML-001 study, a literature review of clinical guidelines and economic models for AML, and advice from two UK clinical oncologists. The structure of the model is described in the submission (Celgene submission, Section 5.5.5, pp. 106–108) and shown in *Figure 9* below.

Figure 9: Structure of the model submitted by Celgene



Source: Celgene submission, Figure 14, p. 107

The model simulates remission, non-remission, relapse/progressive disease and death in AML patients.

The model starts after patients have completed the first cycle of either AZA or CCR (i.e., 4 weeks after the start of treatment). Those patients, who have responded to the treatment by the end of the first cycle and have complete remission (CR) or complete remission with incomplete blood count recovery (CRi), enter the model in “Remission” state; patients with partial remission (PR) or stable disease are placed in “Non-remission” state.

Disease pathways in the subsequent treatment cycles are as follows:

- Patients in the remission state may continue in remission, relapse or die.
- Those in the non-remission state may remain in this state, progress to “Relapse/Progressive disease” state or die.
- Patients in the “Relapse/Progressive disease” state may remain in this state or die.

For every model cycle, the proportion of patients in each health state is estimated using RFS, PFS and OS curves in the following way:

- The proportion of patients in the remission state is based on RFS from AZA-AML-001 study. In the trial, RFS was estimated from the date of first documented CR or CRi to the date of relapse, while OS and PFS were measured from randomisation to event. Therefore, in the model, RFS was adjusted to ensure consistency with OS and PFS.
- The proportion of patients in the non-remission state was estimated from PFS curve for patients who have achieved partial response (PR) or stable disease (SD).
- The proportion of patients in the relapse/progressive disease (PD) state was the difference between the proportions in OS state, and RFS and PFS states.
- The proportion in the “Death” state was a complement of the OS curve.

The model cycle length is 4 weeks which corresponds to one treatment cycle of azacitidine, and is in line with AZA-AML-001 study.

The other key structural assumptions are presented in *Table 17*.

Table 17: Key assumptions in Celgene's economic model

Assumption	Justification
Patients are not eligible for HSCT at any point	The marketing authorisation extension for azacitidine excludes those patients who are eligible for HSCT. Patients in AZA-AML-001 were ineligible for HSCT.
Patients who do not achieve remission in the treatment phase do not subsequently achieve remission	Clinical expert advice. Once off treatment and not in remission, a patient will not achieve remission.
Once in the PD state, patients either remain in PD or die	Clinical expert opinion and previous TAs in similar end-of-life cancers.
There is no treatment switching	Clinical expert opinion. Only a very small percentage of patients at this stage of disease would be fit for a second treatment after failing their first.
In any cycle, patients can only be in one of the health states	Markov model Structure

Key: HSCT, haematopoietic stem cell transplantation; BSC, best supportive care; TAs, technology appraisals; PD, progressive disease.
Source: Celgene submission, Table 54, p. 141

5.2.2 Population

The model population, parameterised from the AZA-AML-001 trial, represents older patients with de-novo or secondary AML with more than 30% bone marrow blasts who were not eligible for hematopoietic stem cell transplantation (HSCT), with intermediate- or poor-risk cytogenetics (NCCN 2009 criteria), Eastern Cooperative Oncology Group performance status (ECOG PS) scores 0–2, and white blood cell count not more than 15×10^9/L.

The starting age of the model population was 75 years and the patient's body surface area (BSA) was 1.8 m², based on the mean age and BSA of AML patients from the AZA-AML-001 trial.

The model allows cross-over adjusted, cross-over unadjusted and censor-at-switch analyses performed for the whole population, IC, LDAC, or BSC patient subpopulations, subpopulations with different cytogenetic risk (intermediate and poor), with and without myelodysplasia-related changes. The following rationale was given for these subgroups (Source: Celgene submission, Section 5.2.1, p. 106):

> *Subgroups on cytogenetic risk and myelodysplasia-related changes were chosen based upon a current unmet need for an effective option for patients presenting with these characteristics and the observed significant (P-values < 0.05) OS benefit of azacitidine over CCR in the AZA-AML-001 trial in these subgroups.*

The results from alternative and subgroup analysis are reported in Section 5.8 of Celgene's submission.

Celgene presented cost-effectiveness results for two subgroups: patients with poor-risk cytogenetics and patients with MDS related changes:

> *Since subsequent-treatment adjustment was not possible for these subgroups, analysis was performed without adjustment.*

5.2.3 Interventions and comparators

The model, proposed by Celgene, estimates cost-effectiveness of AZA compared to conventional care regimens (CCR) comprised of IC+BSC, LDAC+BSC and BSC alone.

Further details regarding the interventions and comparators were given (Source: Celgene submission, Section 5.2.3, pp. 108–109):

> *IC:*
>
> - *Induction therapy – Cytarabine was administered at a dose of 100–200 mg/m²/day via continuous IV infusion for a total of 7 days. Anthracycline was given in combination with cytarabine for 7 days.*
> - *Consolidation - Two consolidation cycles for those who responded to the treatment, followed by BSC. Those who do not respond to induction therapy receive BSC.*
>
> *LDAC: Cytarabine at a dose of 20 mg SC [twice daily] for 10 days, every 28 days, until disease progression or unacceptable toxicity; patients then receive BSC.*
>
> *BSC: Including but is not limited to red cell or whole blood transfusions, fresh frozen plasma transfusions, platelet transfusions, antibiotic and/or antifungal therapy, and nutritional support). This is continued until death.*
>
> *The same BSC is assumed to apply to all patients who have stopped active treatment on AZA, IC or LDAC.*
>
> *Azacitidine is incorporated at a dose of 75 mg/m²/day SC for 7 days every 28 days. In the base case wastage is assumed and vials used are rounded up to the cost of the nearest full vial. Vial sharing is tested in the sensitivity analysis (this also applies to CCR regimens).*

The distribution of patients over IC, LDAC and BSC treatments (18%, 64% and 18%, respectively), modelled in the base case, was derived from the pivotal RCT. It differed substantially from the distribution observed in UK clinical practice and reported in the HMRN registry ([redacted] of patients were treated with IC, [redacted] with LDAC and [redacted] with BSC). To assess the effect of this assumption on the outcome, the manufacturer performed a scenario analysis (Celgene submission, Section 5.8.3, pp.154-156) by calculating a weighted average ICER from individual CCR and AZA arms, with weights equal to the proportions of patients in individual CCR from the HMRN registry.

5.2.4 Perspective, time horizon and discounting

In the model, the perspective on costs was NHS and personal social services perspectives, and the perspective on health effects was direct health effects on patients, in accordance with the NICE reference case.

The model time horizon was 10-years, which was considered the life-time horizon of the patient population in question.

For utilities and costs, the manufacturer used the discount rate of 3.5%, in line with the NICE reference case. In addition, the model allows analyses with the discount rates of 1.5% and 6%.

5.2.5 Treatment effectiveness and extrapolation

Treatment effectiveness was estimated from the AZA-AML-001 trial and post-hoc analyses conducted on the data collected.

The economic model used the following clinical endpoints:

- Overall survival (OS), defined as the time from randomization to death from any cause;
- Relapse-free survival (RFS), the time from first documented CR or CRi to relapse, death from any cause, or loss to follow-up (for the model this was adjusted to time from randomization until relapse or death);
- Progression-free survival (PFS), defined as the time from randomization to death or disease progression (PD) for patients who did not achieve remission (CR or CRi);
- Event-free survival (EFS), defined as the time from randomization to treatment failure, disease progression, relapse after CR or CRi, death from any cause, or loss to follow-up).

EFS was employed to estimate both RFS and PFS, which were obtained (Source: Celgene submission, Section 5.3.1., p. 109):

> *[…] by disaggregating the data into those who did or did not achieve CR or CRi; then, patients with CR or CRi were assessed for death or relapse (i.e., RFS); patients with no CR or CRi were instead assessed for death or disease progression (i.e., PFS).*
>
> *Response status was also used to allocate utilities and disease management costs; in particular, costs for consolidation IC were attributed to patients with CR, CRi, and PR. The cost of BSC was allocated to patients with PD after stopping active treatment.*

In response to an ERG question for clarification regarding the treatment of loss to follow-up, Celgene provided further information:

> *Specific definitions for event free survival (EFS) and relapse free survival (RFS) outcomes are provided below:*
>
> - *Loss to follow-up was treated as an event for EFS outcomes when such loss occurred without documented treatment failure, progression or relapse from*

complete remission (CR)/complete remission with incomplete blood count recovery (Cri) and alive at last contact.

- *Loss to follow-up was treated as an event for RFS outcomes when such loss occurred after documented CR/CRi without relapse from CR/CRi and alive at last contact.*
- *Loss to follow-up was treated as an event for progression free survival (PFS) outcomes when such loss occurred and the variable PDFLAG = 1 (progressive disease (PD) being the best IRC assessed response).*

This was a conservative approach, as the worst case scenario (e.g. progression, relapse or death) is assumed for subjects who are lost to follow up in the context of such a serious disease that requires ongoing medical attention.

The company also referred to definitions in *Table 18* and *Table 19* reproduced from the AML-001 Statistical Analysis Plan document.

Table 18: Censoring rules for event-free survival

Situation	Date of Event or Censoring	Outcome
Withdrawal and no post-baseline response assessments and alive at date of last contact	Date of randomization	Censored
Death without any adequate response assessment	Date of death	Event
Treatment failure, disease progression, relapse after CR/CRi, or death	Earliest of:	Event
	Date of treatment failure	
	Date of disease progression	
	Date of relapse from CR or CRi	
	Date of death	
Lost to follow-up without documented treatment failure, progression, or relapse from CR/CRi and alive at last contact	Date of last response assessment	Event
No treatment failure, progression, or relapse from CR/CRi and not lost to follow-up	Date of last response assessment of CR, CRi, PR, or SD	Censored

Key: CR, complete remission; CRi, complete remission with incomplete blood count recovery; PR, partial remission; SD, stable disease
Source: AML-001 Statistical Analysis Plan Dated Jan 31, 2014

Table 19: Censoring rules for relapse-free survival

Situation	Date of Event or Censoring	Outcome
Relapse or death after CR/CRi	Earliest of:	Event
	Date of relapse from CR or CRi	
	Date of death	
Lost to follow-up after documented CR/CRi without relapse from CR/CRi and alive at last contact	Date of last response assessment	Event
CR/CRi without documented relapse and not lost to follow-up and alive at last contact	Date of last response assessment	Censored

Key: CR, complete remission; CRi, complete remission with incomplete blood count recovery
Source: AML-001 Statistical Analysis Plan Dated Jan 31, 2014

Response rates from the trial (after excluding non-confirmable or non-assessable subjects) used in the model are shown in *Table 20*.

Table 20: Response rates in the AZA-AML-001 trial

Response	Azacitidine response rate	CCR response rate
Remission (CR, CRi)	0.28	0.25
Non-response (PR, SD, PD, TF)	0.72	0.75

Key: CR, complete remission; CRi, morphologic complete remission with incomplete blood count recovery; PD, progressive disease; PR, partial remission' SD, stable disease; TF, treatment failure
Source: Celgene submission, Table 31, p. 109

Several models were fitted to extrapolate overall survival, progression-free survival and relapse-free survival data for each RCT arm:

- Exponential;
- Weibull;
- Gompertz;
- Log-logistic;
- Log-normal.

For each of these models and outcomes treatment effects were also estimated in the form of constant hazard ratios from the coefficient of a binary covariate treatment group indicator included in time to event regressions. Because the log-logistic and log-normal are accelerated failure time models they do not have associated constant hazard ratios, although in the submission (Section 5.3.3, pp. 111, 114, 116, Tables 32–24) the company presents a HR estimate from a Cox proportional hazards model as if it was applicable to the log-logistic and log-normal extrapolation models.

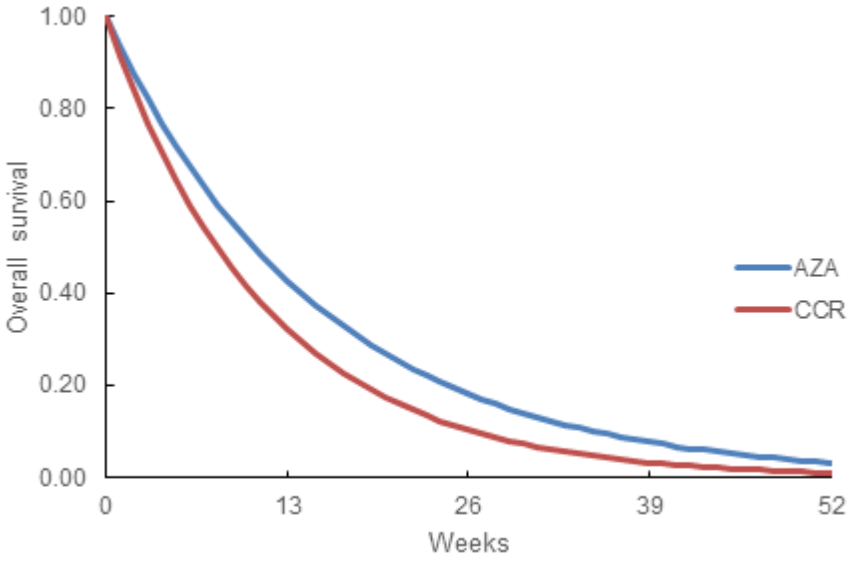
The survival model selection process algorithm recommended by the NICE DSU TSD14 was employed to identify the best fitting curves (Source: Celgene submission, Section 5.3.2, p. 110), with a detailed account on curve fitting also given by the company (Source: Celgene submission, Section 5.3.3, p. 110-116).

For the base case, the manufacturer selected the exponential model for OS, Weibull for RFS and Gompertz for PFS. The company adjusted OS outcomes for treatment switching using a range of different methodological options. PFS and RFS were not adjusted for treatment switching, and Celgene stated that all progression and relapse events in the trial occurred before other censoring events (Celgene submission, Section 5.3.3, p. 116), among which Celgene does not mention treatment switching. In response to the questions for clarification from ERG on why RFS and PFS had not been adjusted for subsequent treatment, the company stated that:

> *This was due to sample size primarily. The instances in which switching preceded the clinical event of interest were few, and the impact of this on the results would be very small.*

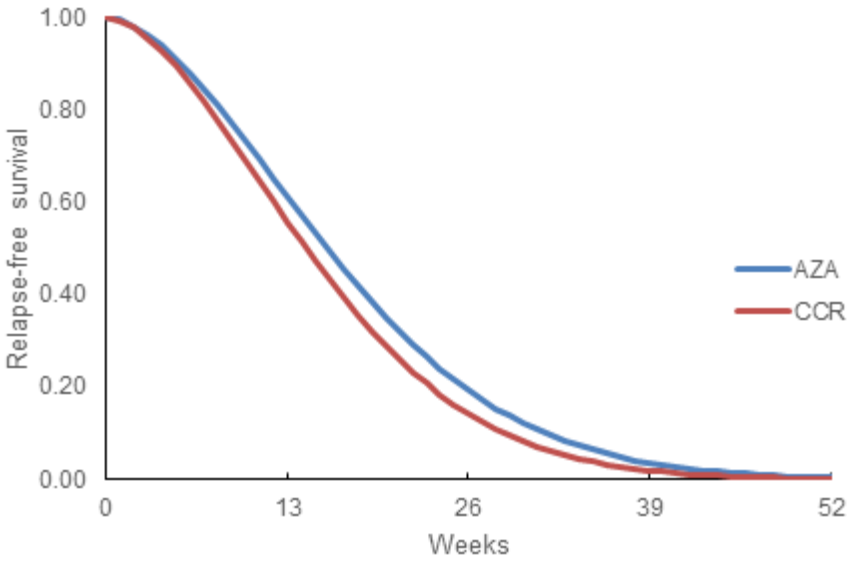
The overall survival, relapse-free survival and progression-free survival used in the model are given in *Figure 10*, *Figure 11* and *Figure 12* respectively.

Figure 10: Overall survival used in the company's submitted model



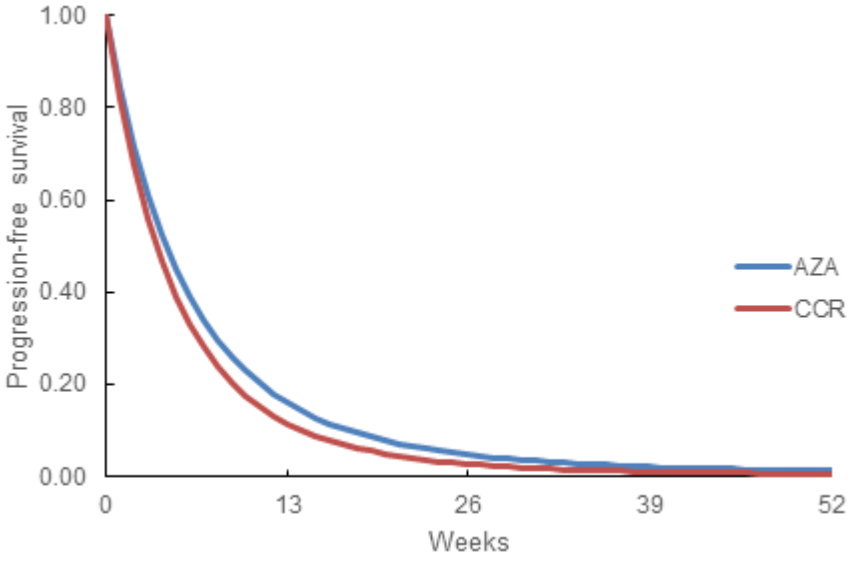
Key: AZA, azacitidine; CCR, conventional care regimens

Figure 11: Relapse-free survival used in the company's submitted model



Key: AZA, azacitidine; CCR, conventional care regimens

Figure 12: Progression-free survival used in the company's submitted model



Key: AZA, azacitidine; CCR, conventional care regimens

5.2.5.1 Treatment switching

The base case OS estimates were based on an adjustment for subsequent treatment use in the CCR arm that censored the data at the time of subsequent treatment initiation and weighted the remaining data by the inverse of the probability of not starting subsequent therapy, which was separately estimated using a logistic regression model. This method is referred to as the inverse probability of censoring weights (IPCW), and was originally developed for the analysis of observational data, applying a similar logic to that used to estimate population statistics from survey data.[57]

The validity of this method hinges on the untestable assumption that information on all determinants of the probability of not using subsequent treatment that are also correlated with OS outcomes is measured in the trial and available to include in the logistic model. This is unlikely to be the case in any practical application, and the question is therefore to what extent any omitted variable bias is likely to be significant. Also, the method may be unstable if few individuals remain uncensored in some patient subgroups as defined by the covariates in the logistic regression (i.e. low effective sample size).

The IPCW method was summarised as follows (Source: Celgene submission, Section 5.3.5, pp. 124–125):

> *Patients who switch treatments are artificially censored at the time of switching, and observations for the remaining patients are weighted to adjust for censored patients. A pooled logistic model is constructed to predict the probabilities of remaining uncensored-by informative censoring (crossover) at each measurement point and must include all baseline or post-randomization variables that predict both treatment switching and outcome. Briefly, the procedure for estimation using IPCW is as follows[52, 57]:*
>
> *1. Panel data are created for the pooled logistic models. The follow-up period is partitioned into intervals based on follow-up measurement points (visit dates). At each measurement point, time-dependent variables that could predict treatment discontinuation, switching, and OS are assessed for all patients.*
>
> *2. The probability of remaining uncensored is calculated. A logistic regression model is fitted to predict participation at each measurement (remaining uncensored) for each subject. The probability of remaining uncensored using baseline risk factors of interest (E) is estimated, as is the probability P of remaining uncensored using both baseline risk factors of interest (E) and time-dependent covariates (Z). The results of this modelling process are summarized to describe the factors associated with participation at each procedure.*
>
> *3. IPCWs are calculated: the inverse probability weight for remaining uncensored (1/P) will consist of the probability for remaining uncensored estimated in step 2, using both covariates E and Z. This inverse probability is stabilized by multiplying it by the probability for remaining uncensored using covariates.*
>
> *4. A standard Cox regression (i.e., in accordance with estimation with no crossover) is fitted for the current outcome using 1/P as weights. The set of covariates E and any other appropriate adjustment covariates for that outcome may also be included in a parametric regression approach. The weighted Cox regression is fitted using stabilized weights (S/P). Standard errors are corrected using sandwich estimation or bootstrapping methods.*
>
> *5. An unweighted version of the Cox regression is fitted for comparison. The same models are fitted as in step 3 but without any sampling weights.*
>
> *Preliminary reviews of the data suggested that subsequent use of azacitidine often closely followed relapse or progression. The model constructed had relatively short time periods (15 days) in order to capture this association. This model was constructed using the status of patients at 15 day time points. The last time period for each study subject usually contained less than 15 days.*

The "numerator" model in the pooled logistic model consisted of baseline factors and the "denominator" model consisted of baseline factors and time-varying covariates. This method provides an estimate of the adjusted HR of survival for the CCR arm in relation to the azacitidine arm but does not generate an estimate of the survival distribution (i.e., does not produce a KM curve). However, a crude estimate of the survival distribution can be obtained by applying the estimated HR to the azacitidine KM curve; this will result in an estimated CCR survival curve with a similar profile (shape) to the azacitidine curve. Similarly, the converse of the HR can be used as an estimate of the median of the adjusted survival distribution for CCR – the actual estimated values are indicative only and will reflect the estimated distribution based on the distribution of the azacitidine KM curve.[58]

Baseline characteristics and time-varying variables were captured during the trial and were used in step 2 of the IPCW method in order to estimate the probabilities of remaining uncensored or having no subsequent use of azacitidine. These variables were assessed by a clinician to establish which factors would be considered relevant and appropriate for use in the crossover analysis models, and whether any of the laboratory variables collected at each visit were relevant for analysis of survival data, either as factors that influence the change in treatment or as factors that could affect the estimate of survival.

Statistical tests were then conducted to assess whether there were any statistically significant differences between CCR patients who switch and CCR patients who do not switch for the list of potential covariates to be included in the model. Means and standard deviations were calculated for numerical variables and counts and percentages for categorical variables for all patients, but also separately for patients who were censored or died. P values were determined using chi-square test for categorical variables and Student's t-test for numerical variables.

The covariates included initially in the model, are presented in [Table 21]. These were summarized by basic summary statistics (number and percentages for categorical variables and means and standard deviations for numerical variables).

Table 21: Covariates used in the company's IPCW analysis

Type of variable	Variable
Non time-varying covariates	
Demographic characteristics	Age at informed consent (continuous)
	Age (<75 years, ≥75 years)
	Sex (male or female)
	Geographic region (North America/Australia, Western Europe/Israel, Eastern Europe, or Asia)
	Race (Asian, Black or African American, White, Native Hawaiian or other Pacific islander, other, n/a)
Clinical characteristics	ECOG performance status at randomization (0–1, 2)
	AML classification (newly diagnosed, histologically confirmed de novo AML; AML secondary to prior myelodysplastic disease not treated with azacitidine, decitabine, or cytarabine; AML secondary to exposure to potentially leukemogenic therapies or agents with the primary malignancy in remission for at least 2 years)
	Time since initial AML diagnosis to randomization (< median; ≥ median) (derived from time since initial diagnosis and date of signed informed consent)
	Baseline comorbidity score
	Prior history of myelodysplastic syndromes (yes or no)
	Cytogenetic risk status (intermediate risk, poor risk)
Study design	Pre-randomization CCR assignment (BSC, low-dose cytarabine, intensive chemotherapy)
	International working group response assessment
Laboratory variables	Percentage bone marrow blasts (continuous) according to central review
Time-varying covariates	
Laboratory variables	WBC count
	Haemoglobin
	Platelet count
	ANC
	RBC transfusion status (independent or dependent)
	Platelet transfusion status (independent or dependent)
Adverse events	Occurrence of a grade 3/4 adverse event since last visit (yes/no)
Other	Time since last visit (in months; included at each visit)

Key: ANC, absolute neutrophil count; ECOG, Eastern Cooperative Oncology Group; RBC, red blood cell; WBC, white blood cell

Source: Celgene submission, Table 37, p. 125

In relation to the statement that "this method provides an estimate of the adjusted HR of survival for the CCR arm in relation to the azacitidine arm but does not generate an estimate of the survival distribution (i.e., does not produce a KM curve)," the ERG notes that this statement is incorrect, because survival curves for both trials arm may be generated from the IPCW analysis. The respective figure is produced by ERG in *Figure 8* (*page 62*) using individual patient data provided by Celgene.

ERG also consulted clinical experts on whether there were any variables missing from *Table 21* that may be considered to predict or explain the use of subsequent therapy and at the same time be associated with survival prognosis. The clinical experts did not suggest any additional variables.

In arriving at their preferred base case method the company also considered other approaches, including not adjusting for treatment switching, censoring at switch, and the Rank Preserving Structural Failure Time Model (RPSFTM) and the Iterative Parameter Estimation (IPE). Celgene considered the RPSTFM and IPE approaches as involving a fundamental assumption (i.e., that the treatment effect of azacitidine is the same for patients who received it as initially randomly allocated therapy as for patients who received it as subsequent treatment) that is unlikely to be valid. The company stated (Source: Celgene submission, Section 5.3.5, p. 123):

> *These limits notwithstanding, the IPCW method and results are stronger than the alternatives. The use of RPSFT and IPE methods has an underlying assumption of a common treatment effect for patients who started treatment with azacitidine and for those who switched to azacitidine. This assumption does not hold in this case: differences in prognosis between the two groups are likely to lead to a different benefit from delayed versus immediate treatment; CCR itself, particularly LDAC and IC, is also an active treatment, so the prognosis of a patient switching from CCR will not be the same as for a patient receiving azacitidine from the start of the study.*
>
> *The similarity of the results for the IPE and RPSFT analyses is likely because of violation of the key assumption, that treatment benefit in terms of OS is the same regardless of whether a patient began on azacitidine or switched; this assumption is hard to justify given the prognosis for with AML in the trial.*

In addition, Celgene reported the attempt to use of two-stage methods,[51] which could not be implemented due to insufficient numbers of subsequent treatment users in the CCR arm and recorded data (Source: Celgene appendices to submission, Appendix 11, Section 4.7).

Similarly to the IPCW method, the censor-at-switch method dropped the data after the start of azacitidine in the CCR trial arm, but unlike IPCW it did not apply any adjustment to the remaining data for the differences in probability of censoring (azacitidine) across CCR arm patients. Thus, to the extent that the estimates from the two methods differ (provided IPCW is validly in this application), would suggest that azacitidine was not given to CCR patients at random (as one would expect). As it transpires, Celgene reported hazard ratio estimates from Cox proportional hazards models of [redacted] for both the IPCW and censor-at-switch methods, suggesting that adjusting for non-random subsequent therapy use is not important in this case (see also the critique of treatment switching in the clinical effectiveness *Section 4.3*).

5.2.5.2 Summary of methods of effectiveness estimation and extrapolation

In response to ERG's request to confirm the methods used to calculate different survival curves in the model, and whether the curves for relapse-free survival and progression-free survival were fitted to azacitidine (AZA) and conventional care regimen (CCR) patients with a proportional-hazards azacitidine treatment variable, or if these were fitted only to CCR patients, Celgene provided the following information:

> *The following methods were used to calculate different survival curves. The curves for RFS and PFS were fitted only to CCR patients.*

Table 22: Methods used to calculate survival curves in the model submitted by Celgene

Arm	AZA	CCR
Overall survival		
Underlying data	OS from AZA	OS from AZA
Curve fitting	Exponential	Exponential
Adjustments	—	HR of [redacted] from IPCW method (inverse HR)
Relapse-free survival		
Underlying data	EFS for CCR patients achieving CR or CRi	EFS for CCR patients achieving CR or CRi
Curve fitting	Weibull	Weibull
Adjustments	HR of 0.84 from curve fitting	—
Progression-free survival		
Underlying data	EFS for CCR patients not achieving CR or CRi	EFS for CCR patients not achieving CR or CRi
Curve fitting	Gompertz	Gompertz
Adjustments	HR of 0.85 from curve fitting	—

Key: AZA, azacitidine; CCR, conventional care regimen; CR, complete remission; CRi, complete remission with incomplete blood count recovery; EFS, event-free survival; HR, hazard ratio; IPCW, inverse probability of censoring weights

5.2.6 Health-related quality of life

Health effects were measured in QALYs in accordance with the NICE reference case.

Utilities were estimated from response status. They were mapped from trial-based EORTC QLQ-C30 data using published algorithms (Source: Celgene submission, Section 5.4, p. 127). Two mapping algorithms were incorporated in the model, one reported by Proskorovsky et al. 2014,[59] which was used for the base case, and the other by McKenzie and Van der Pol, 2009,[60] used for a scenario analysis. The algorithms are presented in Celgene's Submission, Table 40 (p. 128) and the corresponding utility values are shown in *Table 23*.

Table 23: Summary of utility values for Celgene's economic evaluation

Health state	Calculation method	
	Proskorovsky et al. 2014[59] *(base case)*	*McKenzie and Van der Pol 2009*[60] *(scenario analysis)*
Remission (CR/CRi)	0.7707	0.7400
Non-remission (PR, SD)	0.7160	0.6574
Post-progression/relapse (PD)	0.6233	0.5680
Grade 3+ AEs	−0.0240	−0.0207

Key: AE, adverse event; CR, complete remission; CRi, morphologic complete remission with incomplete blood count recovery; PD, progressive disease; PR, partial remission' SD, stable disease
Source: Celgene submission, Table 41, p. 129

The model accounts for disutility associated with overall grade 3 or above treatment-emergent adverse events (TEAEs) (Source: Celgene submission, Table 41, p. 129). The EORTC QLQ-C30 data used to map the EQ-5D utility for TEAEs are from patients who were hospitalised with and without grade 3 or higher TEAEs in the AZA-AML-001 trial.

Adverse event-related QALYs were calculated by (Source: Celgene submission, Section 5.4.3, p. 129):

> *[Multiplying the probability of at least one TEAE occurring] by its duration in days, then multiplying the result by the day equivalent of the HSUV.*

5.2.7 Resources and costs

5.2.7.1 Drug acquisition

Drug utilization was estimated directly from the AZA-AML-001 trial.

Total drug use per cycle per patient (*Table 24*) was calculated by multiplying the number of vials per day per patient, the cost of one vial (*Table 25*) and the mean number of treatment days per cycle (*Table 26*).

The number of vials per day per patient was based on the average daily dose (*Table 24*) and the assumption on vial sharing, i.e., no wastage, full wastage or wastage with 30% tolerance.

For azacitidine, IC and LDAC treatments, the average daily dose was estimated from the average daily dose in mg/m² (*Table 24*) and the average body surface area (BSA) of 1.80 m². However, according to the CSR, the daily dose for LDAC treatment should have been estimated in mg/day (further details are provided in *Section 5.3.8*, *page 122*).

Table 24: Drug utilisation per cycle (4 weeks) for no wastage scenario

Treatment	Medications	Daily dose (mg/m^2)	Days per cycle	Total Dose (mg) per cycle
Azacitidine	Azacitidine	[redacted]	[redacted]	[redacted]
IC, induction	Cytarabine	122.20	7.10	1,561.72
	Daunorubicin[a]	49.70	3.00	268.38
	Idarubicin[a]	11.00	3.00	59.40
IC, consolidation	Cytarabine	120.20	5.00	1,081.80
	Daunorubicin[a]	49.40	2.00	177.84
	Idarubicin[a]	10.70	2.00	38.52
LDAC	Cytarabine	84.05	10.22	696.65

Key: IC, intensive chemotherapy; LDAC, low-dose cytarabine
Notes: a, Use of anthracycline in the trial comprised 50% idarubicin and 50% daunorubicin
Source: Celgene submission, Table 42, p. 131

Full wastage (i.e., no vial sharing) was assumed in the base case, and alternative scenarios of no wastage and wastage with 30% tolerance (i.e., vial sharing assumed in 30% of cases) were explored in sensitivity analyses.

For drugs with several vial or pack sizes, vial size selection was on the basis of the largest available size, rather than smaller vials as required to minimize vial wastage. The number of vials for each drug was not reported in the company's submission, but was available from their executable model.

Drug acquisition unit costs, presented in *Table 25*, were estimated from British National Formulary (BNF).[61]

Table 25: Drug acquisition unit costs

Drug name	Vial or pack	mg per vial or pack	Price (£) per vial/pack
Azacitidine	100 mg vial	100	[redacted]
Cytarabine (non-proprietary)	20 mg/mL; 5 mL vial or 100 mg/mL; 1 mL vial	100	4.95[a]
	20 mg/mL; 25 mL vial or 100 mg/mL; 5 mL vial	500	19.75[a]
	100 mg/mL; 10 mL vial	1000	39
	100 mg/mL; 20 mL vial	2000	77.5
Daunorubicin (non-proprietary)	20 mg vial	20	55
Idarubicin (Zavedos®)	5 mg vial	5	87.36
	10 mg vial	10	174.72

Notes: a, Average of two prices
Sources: Celgene submission, Table 43, p. 132; BNF 2015

Table 26: Drug acquisition cost per cycle

Treatment		Total drug cost per cycle per patient (£)		
		No wastage	*Wastage*	*Wastage with 30% tolerance*
Azacitidine		[redacted]	[redacted]	[redacted]
IC, induction	Cytarabine	£77	£105	£77
	Daunorubicin[a]	£738	£825	£738
	Idarubicin[a]	£1,038	£1,048	£1,038
IC, consolidation	Cytarabine	£54	£75	£54
	Daunorubicin[a]	£550	£489	£489
	Idarubicin[a]	£673	£699	£673
LDAC	Cytarabine	£34	£48	£34

Key: IC, intensive chemotherapy; LDAC, low-dose cytarabine
Notes: a, Average of two prices
Source: Celgene submission, Table 44, p. 132

The total cost of drug acquisition was estimated from the per-cycle cost of drugs and setting the maximum number of treatment cycles equal to the average number of treatment cycles shown in *Table 27*.

Table 27: Mean number of treatment cycles in the AZA-AML-001 trial

Treatment		Mean number of cycles per patient
Azacitidine		8.80
IC, induction	Cytarabine	1.00
	Daunorubicin[a]	1.00
	Idarubicin[a]	1.00
IC, consolidation	Cytarabine	1.00
	Daunorubicin[a]	1.00
	Idarubicin[a]	1.00
LDAC	Cytarabine	5.21
BSC		3.60

Key: BSC, best supportive care; IC, intensive chemotherapy; LDAC, low-dose cytarabine
Notes: a, 1:1 ratio was assumed for patients on daunorubicin and idarubicin
Source: Celgene submission, Table 45, p. 133

5.2.7.2 Drug administration

Health-care resource use (HCRU) for each model state is reported in Celgene's submission, Section 5.5.3 (p. 133).

Unit costs for health professionals, used in the model, are from Personal Social Service Research Unit (PSSRU) and NHS reference costs (*Table 28*).

Table 28: Unit costs for each item of healthcare resource use

Staff type	Unit costs available 2013/2014 (costs including qualifications given in brackets)	Cost per minute (per day for inpatient stay)
CNS Haematologist	Nurse advanced (includes lead specialist, clinical nurse specialist, senior specialist). £51 (£58) per hour; £80 (£90) per hour client contact cost	£1.33
Consultant	Consultant: medical, £101 (£140) per contract hour	£1.68
Day Care Nurse	Nurse, day ward (includes staff nurse, registered nurse, registered practitioner), £34 (£41) per hour; £84 (£100) per hour of patient contact	£1.40
Day Care Specialist registrar	Registrar group, £40 (£60) per hour (48 hour week); £34 (£51) per hour (56 hour week); £48 (£71) per hour (40 hour week)	£0.80
District Nurse	Community nurse (includes district nursing sister, district nurse), £43 (£50) per hour; £57 (£66) per hour of patient-related work.	£0.95
Doctor	Associate specialist, £97 (£124) per hour (40 hour week). An associate specialist is a doctor who has trained and gained experience in a medical or surgical specialty but has not become a consultant.	£1.62
Jnr. Doctor	Foundation house officer 2, £29 (£41) per hour (48 hour week); £25 (£35) per hour (56 hour week); £35 (£49) per hour (40 hour week)	£0.58
Pharmacist	Hospital pharmacist, £42 (£48) per hour; £84 (£96) per cost of direct clinical patient time (includes travel); £60 (£68) per cost of patient-related activities.	£1.40
Oncology nurse	Nurse team leader (includes deputy ward/unit manager, ward team leader, senior staff nurse), £42 (£48) per hour; £104 (£120) per hour of patient contact	£1.73
Inpatient stay for IC (cost/day)	Average of "Elective Inpatients - Excess Bed Days", "Non-Elective Inpatients - (Long Stay) Excess Bed Days", "Day Case", "Non-elective Inpatients - Short Stay", "Regular Day or Night Admissions"	£714.64

Key: CNS, clinical nurse specialist; SpR, specialist registrar; PSSRU, Personal Social Services Research Unit; IC, intensive chemotherapy

Sources: Celgene submission, Table 48, p. 136; PSSRU Unit costs of health and social care 2014[62]; NHS Reference costs

The frequency (*Table 29*) and mean time in minutes of health professionals involved during different health states (*Table 30*) were estimated from a healthcare resource use questionnaire developed by Celgene (Celgene's appendices to submission, Appendix 12).

Table 29: Healthcare resource use (frequency per cycle) for each health state

Healthcare resouce	Induction/pre-response		Remission		Stable disease		Progressive disease	
	AZA	*CCR*	*AZA*	*CCR*	*AZA*	*CCR*	*AZA*	*CCR*
CNS Haematologist	2.77	2.38	1.66	0.87	2.08	2.37	2.03	2.62
Consultant	2.58	3.52	0.92	1.13	1.29	1.66	2.03	1.60
Day Care Nurse	7.75	2.35	5.54	0.95	6.00	3.41	3.69	3.47
Day Care SpR	1.66	5.16	1.11	2.07	1.66	2.85	2.95	2.95
District Nurse	0.62	5.39	0.31	5.61	0.62	6.33	0.62	0.59
Doctor	0.85	4.95	1.23	2.21	1.54	3.04	0.92	0.88
Jnr. Doctor	0.23	17.11	0.62	21.75	2.54	17.31	2.77	2.64
Pharmacist	2.77	3.09	2.77	1.78	2.95	1.37	0.31	0.42
Oncology nurse	0.62	2.17	0.31	0.05	0.62	0.50	0.62	0.59
Inpatient day	3.16	13.91	0.25	0.90	2.30	9.20	1.73	2.61

Key: AZA, azacitidine; CCR, conventional care regimens; CNS, clinical nurse specialist; SpR, specialist registrar
Source: Celgene submission, Table 46, p. 134

Table 30: Healthcare resource use (mean time in minutes per frequency) for each health state

Healthcare resource	Induction/pre-response		Remission		Stable disease		Progressive disease	
	AZA	*CCR*	*AZA*	*CCR*	*AZA*	*CCR*	*AZA*	*CCR*
CNS Haematologist	34.20	25.51	26.40	25.85	33.00	37.23	24.40	34.20
Consultant	25.60	20.33	20.80	16.29	24.00	16.97	20.80	25.60
Day Care Nurse	40.72	22.99	18.40	3.91	26.83	27.55	33.00	40.72
Day Care SpR	22.00	19.82	22.00	17.79	22.00	19.08	22.00	22.00
District Nurse	15.00	13.87	15.00	4.26	15.00	17.31	15.00	15.00
Doctor	12.67	13.00	12.67	9.79	12.67	12.10	9.00	12.67
Jnr. Doctor	9.00	16.01	15.00	0.94	20.00	10.76	12.67	9.00
Pharmacist	13.50	25.64	13.50	0.71	13.50	19.68	6.00	13.50
Oncology nurse	6.00	11.68	4.00	0.00	6.00	4.55	6.00	6.00
Inpatient day	1,440	1,440	1,440	1,440	1,440	1,440	1,440	1,440

Key: AZA, azacitidine; CCR, conventional care regimens; CNS, clinical nurse specialist; SpR, specialist registrar
Source: Celgene submission, Table 47, p. 135

Monitoring and testing requirements for different disease states are captured in the same questionnaire (Celgene appendices to submission, Appendix 12). The number of drug monitoring tests per cycle is shown in *Table 31*, and the unit costs in *Table 32*.

Table 31: Healthcare resource use (number of tests per cycle) for drug monitoring tests

Monitoring test	Induction/pre-response		Remission		Stable disease		Progressive disease	
	AZA	*CCR*	*AZA*	*CCR*	*AZA*	*CCR*	*AZA*	*CCR*
Bone marrow aspirates	0.92	1.25	0.15	0.21	0.42	0.35	0.15	0.16
Bone marrow biopsies	0.50	0.41	0.00	0.00	0.04	0.08	0.00	0.03
Peripheral blood smears	1.08	1.01	0.77	0.59	0.77	0.83	0.77	0.74
Blood tests	9.23	13.29	1.85	3.53	6.54	7.82	7.23	8.33
DNA and RNA extractions for molecular testing	0.92	1.24	0.15	0.15	0.15	0.20	0.15	0.15
Extractions for cytogenetic testing	0.92	0.80	0.15	0.16	0.15	0.19	0.15	0.13
Serum blood chemistry	8.46	12.00	1.69	3.53	6.38	7.72	6.92	7.74

Key: AZA, azacitidine; CCR, conventional care regimens; DNA, deoxyribonucleic acid; RNA, ribonucleic acid
Source: Celgene submission, Table 49, p. 137

Table 32: Unit costs for drug monitoring tests

Laboratory and disease monitoring tests	HRG (Description)	Cost per test
Bone marrow aspirates	DAPS04 (Clinical Biochemistry, National average cost)	£1.18
Bone marrow biopsies	DAPS04 (Clinical Biochemistry, National average cost)	£1.18
Peripheral blood smears	DAPS05 (Haematology, National average cost)	£3.00
Blood tests	DAPS05 (Haematology, National average cost)	£3.00
DNA and RNA extractions for molecular testing	DAPS04 (Clinical Biochemistry, National average cost)	£1.18
Extractions for cytogenetic testing	DAPS01 (Cytology, National average cost)	£7.77
Serum blood chemistry	DAPS04 (Clinical Biochemistry, National average cost)	£1.18

Key: DNA, deoxyribonucleic acid; HRG, healthcare resource group; RNA, ribonucleic acid
Source: Celgene submission, Table 50, p. 137; NHS reference costs 2013–14

Red blood cell and platelet transfusions were also included in the model (*Table 33*).

Table 33: Unit cost and resource use (number of transfusions per cycle) of transfusions

Transfusion type	Induction/pre-response		Remission		Stable disease		Progressive disease		Unit cost (per transfusion)
	AZA	*CCR*	*AZA*	*CCR*	*AZA*	*CCR*	*AZA*	*CCR*	
Red blood cells	3.62	3.40	0.15	0.72	3.00	3.05	4.55	4.78	£121.85[63]
Platelets	4.54	3.63	0.15	0.48	3.92	3.46	5.70	5.85	£193.15[64]

Key: AZA, azacitidine; CCR, conventional care regimens
Source: Celgene submission, Table 51, p. 138

5.2.7.3 Adverse events

The cost of managing AEs was calculated as a cost per patient, based on the arithmetic average cost for managing grade 3 or 4 TEAEs in the AZA-AML-001 trial such as anaemia, neutropenia, febrile neutropenia, thrombocytopenia, pneumonia, and worsening AML (not qualifying as a progression or relapse event); the cost are shown in *Table 34*.

Table 34: Costs of managing adverse events (≥ grade 3)

Adverse Event	Cost per inpatient episode	Source
Anaemia Neutropenia Febrile neutropenia	£341.69	SA08J (Other Haematological or Splenic Disorders, with CC Score 0-2)[a]
Thrombocytopenia	£316.46	SA12K (Thrombocytopenia with CC Score 0-1)[a]
Pneumonia	£143.64	WF01A Service Code 300 (Consultant-led outpatient attendance, General Medicine, Non-Admitted Face to Face Attendance, Follow-up)[b]
Acute myeloid leukaemia	£377.01	SA25M (Acute Myeloid Leukaemia with CC Score 0-1)[a]
Grade ≥ 3 TEAEs	£310.36	Average

Key: CC, complications and comorbidities; TEAE, treatment-emergent adverse event
Notes: a, Unit day case cost; b, National unit cost
Sources: Celgene submission, Table 52, p. 139; National schedule of reference costs 2013-14

5.2.8 Cost-effectiveness results

Base-case results are shown in *Table 35*. Estimated cost per QALY reported by Celgene is £20,648.

Table 35: Base case results of the company's model

Arm	Total			Incremental			ICER (cost per QALY)
	Costs	LYG	QALYs	Costs	LYG	QALYs	
CCR	£40,608	0.9041	0.6365	—	—	—	—
Azacitidine	[redacted]	1.1820	[redacted]	[redacted]	0.2779	[redacted]	£20,648

Key: ICER, incremental cost-effectiveness ratio; LYG, life years gained; QALYs, quality-adjusted life years; CCR; conventional care regimens

Disaggregated results are shown in *Table 36*, *Table 37* and *Table 38*.

Table 36: Health outcomes (QALYs) by health state in the company's model

Health state	Azacitidine	CCR	Incremental QALYs	Contribution to total incremental QALYs
RFS	[redacted]	0.2312	[redacted]	[redacted]
PFS	[redacted]	0.2725	[redacted]	[redacted]
PD	[redacted]	0.1328	[redacted]	[redacted]
Total	[redacted]	0.6365	[redacted]	[redacted]

Key: CCR, conventional care regimens; PD, progressive disease; PFS, progression-free survival; QALY, quality-adjusted life year; RFS, relapse-free survival
Notes: Figures may not add due to rounding

Table 37: Costs by health state in the company's model

Health state	Azacitidine	CCR	Incremental costs	Contribution to total incremental costs
RFS	[redacted]	£6,503	[redacted]	[redacted]
PFS	[redacted]	£22,235	[redacted]	[redacted]
PD	[redacted]	£6,260	[redacted]	[redacted]
Terminal care	[redacted]	£5,609	[redacted]	[redacted]
Total	[redacted]	£40,608	[redacted]	[redacted]

Key: CCR, conventional care regimens; PD, progressive disease; PFS, progression-free survival; RFS, relapse-free survival
Notes: Figures may not add due to rounding

Table 38: Costs by component in the company's model

Cost component	Azacitidine	CCR	Incremental costs	Contribution to total incremental costs
Drug acquisition	[redacted]	£370	[redacted]	[redacted]
Drug administration	[redacted]	£23,316	[redacted]	[redacted]
Tests to monitor disease	[redacted]	£157	[redacted]	[redacted]
Transfusion	[redacted]	£4,624	[redacted]	[redacted]
Management of AEs	[redacted]	£269	[redacted]	[redacted]
BSC/Monitoring costs	[redacted]	£6,260	[redacted]	[redacted]
Terminal care	[redacted]	£5,609	[redacted]	[redacted]
Total	[redacted]	£40,608	[redacted]	[redacted]

Key: AEs, adverse events; BSC, best supportive care; CCR, conventional care regimens; PD, progressive disease; PFS, progression-free survival; RFS, relapse-free survival
Notes: Figures may not add due to rounding

5.2.8.1 Subgroup analyses

Celgene presented cost-effectiveness results for two subgroups: patients with poor-risk cytogenetics (*Table 39*) and patients with MDS related changes (*Table 40*). For these subgroups, analysis was performed without adjustment for subsequent active treatment.

Table 39: Results for patients with poor-risk cytogenetics

Arm	Total			Incremental			ICER (cost per QALY)
	Costs	LYG	QALYs	Costs	LYG	QALYs	
CCR	£46,683	0.6607	0.4567	—	—	—	—
Azacitidine	[redacted]	1.1855	[redacted]	[redacted]	0.5248	[redacted]	£20,227

Key: ICER, incremental cost-effectiveness ratio; LYG, life years gained; QALYs, quality-adjusted life years; CCR; conventional care regimens
Source: Celgene submission, Table 63, p. 158

Table 40: Results for patients with MDS related changes

Arm	Total			Incremental			ICER (cost per QALY)
	Costs	LYG	QALYs	Costs	LYG	QALYs	
CCR	£50,098	0.9459	0.6583	—	—	—	—
Azacitidine	[redacted]	1.4050	[redacted]	[redacted]	0.4591	[redacted]	£19,175

Key: ICER, incremental cost-effectiveness ratio; LYG, life years gained; QALYs, quality-adjusted life years; CCR; conventional care regimens
Source: Celgene submission, Table 64, p. 158

5.2.9 Sensitivity analyses

Celgene performed deterministic and probabilistic sensitivity analyses to estimate the effect of uncertainty in model parameters on the ICER.

5.2.9.1 Univariate sensitivity analyses

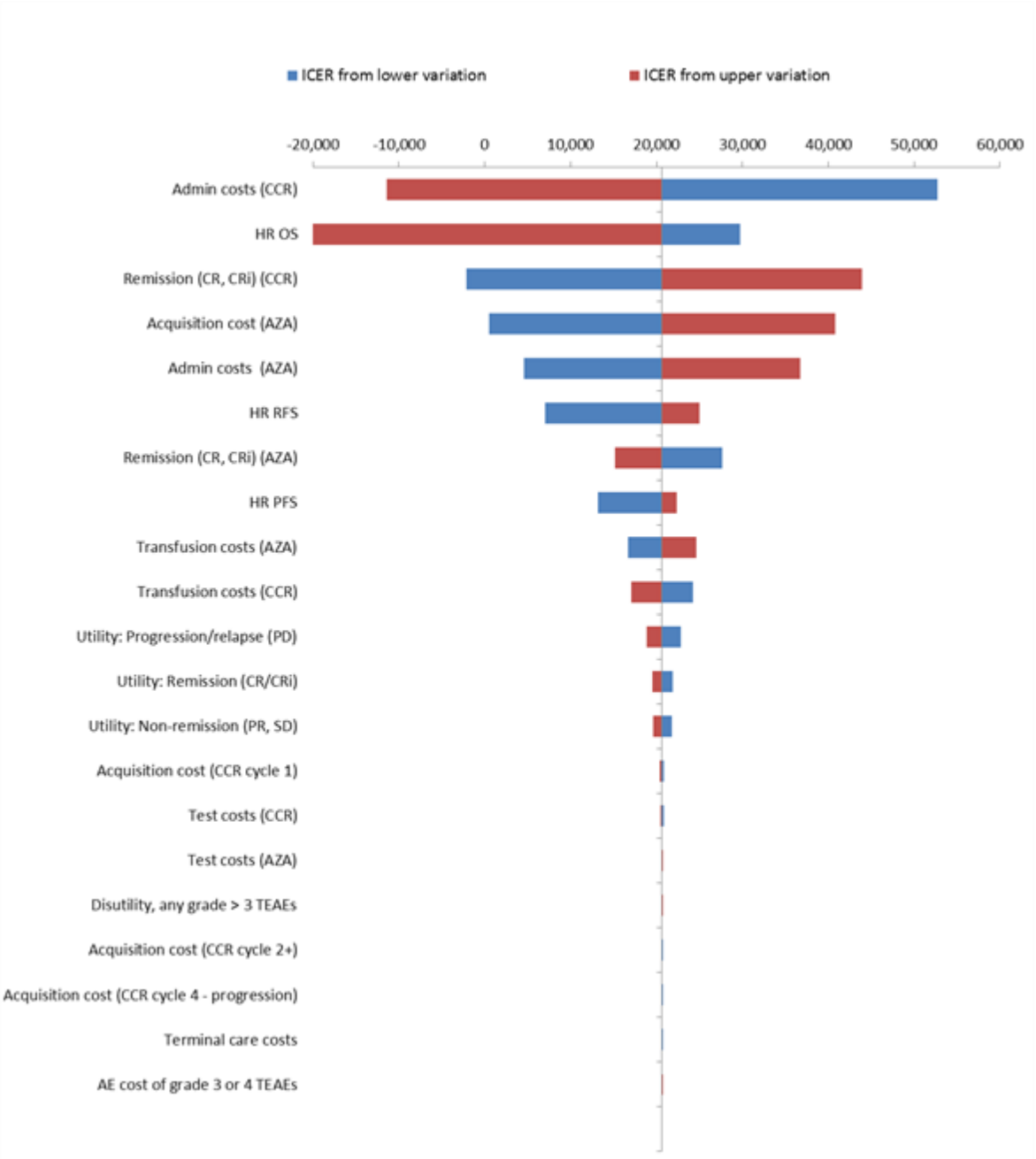
One-way SA is a form of deterministic sensitivity analysis in which one parameter value is varied while keeping all other parameter values constant, to investigate the impact of individual parameters on the base case ICER.

In the model, the base case values of the following parameters were varied by ±20% or around a confidence interval (for HRs) to evaluate this impact:

- Drug utilization costs;
- Drug administration costs;
- Drug monitoring cost (transfusion and tests);
- BSC/palliative care costs;
- HRs;
- Safety;
- Response rate;
- Health state utility values.

A tornado diagram demonstrating the results of the univariate sensitivity analyses is shown in *Figure 13*.

Figure 13: Tornado diagram of company's deterministic sensitivity analyses



Key: AZA, azacitidine; CCR, conventional care regimens; CR, complete remission; CRi, complete remission with incomplete blood count recovery; HR, hazard ratio; ICER, incremental cost-effectiveness ratio; OS, overall survival; PFS, progression-free survival; PD, progressive disease; RFS, relapse-free survival; SD, stable disease; TEAE, treatment-emergent adverse events

Source: Celgene submission, Figure 34, p. 152

5.2.9.2 Probabilistic sensitivity analysis

The manufacturer performed second-order Monte Carlo simulations by randomly drawing from all predefined parameter distributions simultaneously and computed incremental costs and health outcomes for the random variates. The results were plotted as X-Y scatter plot and cost-effectiveness acceptability curves where the "willingness to pay" is plotted against the proportion of runs that resulted in incremental cost-effectiveness ratios below this

willingness to pay. The parameter distributions used for the probabilistic sensitivity analysis are reported in *Table 41*.

Table 41: Distributions used in the company's probabilistic sensitivity analysis

Beta distribution	Gamma distribution
• Response rate • HSUVs • HSUVs for adverse events • Incidence of adverse events	• Patients' weights and heights • Drug usage and number of treatment cycles • Healthcare resource use

Key: HSUV, health state utility value
Source: Celgene submission, Table 59, p. 148

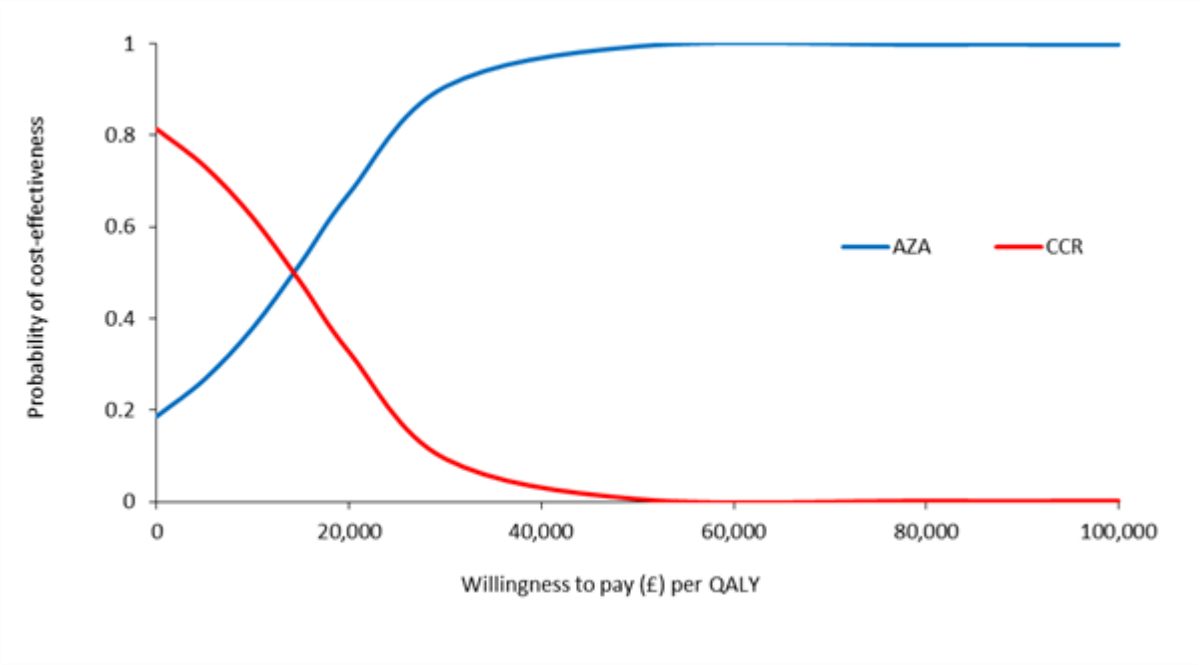
Point estimates and 95% uncertainty intervals from PSA are reported in *Table 42*. Cost-effectiveness acceptability curves are shown in Figure 14, and show that at cost-effectiveness thresholds of £20,000, £30,000 and £50,000 per QALY, azacitidine was cost-effective versus CCR in 69.9%, 90.8% and 99.6% of iterations respectively.

Table 42: Results of the company's submitted probabilistic sensitivity analysis

Arm	Total (95% CI)			Incremental			ICER (cost per QALY)
	Costs	LYG	QALYs	Costs	LYG	QALYs	
CCR	£41,429 (£34,562, £49,698)	0.9073 (0.6970, 1.1358)	0.6386 (0.5047, 0.7924)	—	—	—	—
AZA	[redacted]	1.1824 (1.0337, 1.3468)	[redacted]	[redacted]	0.2751	[redacted]	£17,423

Key: AZA, azacitidine; CCR, conventional care regimens; CI, confidence interval; ICER, incremental cost-effectiveness ratio; LYG, life-years gained; QALY, quality-adjusted life year
Source: Celgene submission, Table 60, p. 149

Figure 14: Cost-effectiveness acceptability curves in the company's probabilistic sensitivity analysis



Key: AZA, azacitidine; CCR, conventional care regimens; QALY, quality-adjusted life year
Source: Celgene submission, Figure 33, p. 151

5.2.9.3 Scenario analyses

Celgene presented the results of the following scenario analyses (Source: Celgene submission, section 5.8.3, p. 154):

> *KM curves for RFS, PFS and OS*
>
> *OS data unadjusted for treatment-switching*
>
> *OS using the censor at switch population*
>
> *EQ-5D based on the mapping algorithm from McKenzie et al*
>
> *Vial Sharing*
>
> *Vial sharing in 30% of cases*
>
> *1 year and 5 year time horizons*
>
> *Discount rate at 1.5% and 6%*
>
> *Individual treatment arms with adjustment for subsequent therapies*
>
> *Individual treatment arms without adjustment for subsequent therapies*

Celgene also conducted a scenario analysis in which the proportion of patients assigned to each CCR arm was estimated from HMRN registry data (IC, [redacted]; LDAC, [redacted]; BSC, [redacted]). A weighted average ICER was calculated by multiplying the total costs and QALYs from the individual CCR and azacitidine arms (azacitidine results from individual arms not CCR population) by these proportions and then summing the resulting totals.

The results of the scenario analyses are shown in *Table 43*.

Table 43: Results of the company's scenario analyses

Scenario		Incremental			ICER (cost per QALY)[a]
		Costs	LYG	QALYs	
Base case		[redacted]	0.2779	[redacted]	£20,648
KM curves for RFS, PFS and OS		[redacted]	0.1485	[redacted]	£32,393
OS data unadjusted for treatment-switching		[redacted]	0.3630	[redacted]	£11,537
OS using the censor at switch population		[redacted]	0.8309	[redacted]	£10,397
EQ-5D based on the mapping algorithm from McKenzie et al.[60]		[redacted]	0.2779	[redacted]	£22,243
Vial sharing		[redacted]	0.2779	[redacted]	-£13,300
Vial sharing in 30% of cases		[redacted]	0.2779	[redacted]	-£9,323
Time Horizon	1 year	[redacted]	0.0791	[redacted]	£30,305
	5 year	[redacted]	0.2673	[redacted]	£20,860
Discount Rate	1.50%	[redacted]	0.2861	[redacted]	£20,604
	6%	[redacted]	0.2685	[redacted]	£20,704
Individual treatment arms with adjustment for subsequent therapies	IC	[redacted]	0.3759	[redacted]	-£52,184
	LDAC	[redacted]	0.2729	[redacted]	£25,136
	BSC	[redacted]	0.2095	[redacted]	-£169,672
Individual treatment arms without adjustment for subsequent therapies	IC	[redacted]	0.2449	[redacted]	-£85,266
	LDAC	[redacted]	0.2600	[redacted]	£41,671
	BSC	[redacted]	0.3386	[redacted]	-£50,300
Use of individual treatment arm proportions from the HMRN registry with adjustment for subsequent therapies		[redacted]	0.2665	[redacted]	-£57,756
Use of individual treatment arm proportions from the HMRN registry without adjustment for subsequent therapies		[redacted]	0.2874	[redacted]	-£20,218

Key: BSC, best supportive care; EQ-5D, EuroQol Five Dimensions; HMRN, Haematological Malignancy Research Network; IC, intensive chemotherapy; ICER, incremental cost-effectiveness ratio; KM, Kaplan-Meier; LDAC, low dose cytarabine; LYG, life years gained; OS, overall survival; PFS, progression-free survival; QALY, quality-adjusted life year; RFS, relapse-free survival
Note: a, Negative ICER indicates that azacitidine is dominant
Source: Celgene submission, Table 62, p. 156

5.2.10 Model validation and face validity check

The manufacturer performed the following model validation and verification.

Model structure and assumptions were assessed at four levels (Source: Celgene submission, Section 5.9.1, p. 159):

> *An internal clinical validation was performed by PRMA Consulting's Senior Medical Director, Professor Deborah Saltman.*

> *PRMA Consulting's senior management and expert health economists performed an internal validation.*
>
> *The validity of the model was confirmed by Professor Stephen Palmer, an external technical advisor with extensive experience of NICE HTAs.*
>
> *Externally, two UK clinical oncologists validated the model structure and key assumptions; one of them also validated HCRU inputs (types of HCRU involved) and model outputs on effectiveness.*
>
> *The model was also reviewed by the Celgene team.*

Technical validity was ensured by performing model checks reported in *Table 44*.

Table 44: Checklist used to check the model inputs and results

Check	Purpose
Set discount rate to 0	To confirm that discounted and non-discounted results are equal
Set main HSUVs to 0	To confirm that QALYs are zero, or can be explained by utility decrements associated with adverse events
Set all HSUVs to 1	To confirm that LYs are equal to QALYs, or that any difference can be explained by utility decrements associated with adverse events
Set drug costs to 0	To confirm that drug costs are zero
Set admin costs to 0	To confirm that administration costs are zero
Set all non-drug costs to 0	To confirm that non-drug costs are zero
Manually confirm tornado diagram calculations by changing user-altered cells	To confirm that that tornado diagram calculations are correct

Key: HSUV, health state utility value; LY, life-year; QALY, quality-adjusted life-year
Source: Celgene submission, Table 65, p. 159

The OS for patients in CCR arm, predicted by the model, has been compared to UK real world data from the HMRN registry. *Figure 15* provides a comparison for all patients and *Figure 16* for those patients with poor-risk cytogenetics. Celgene's submission states (Source: Celgene Submission, Section 5.9.3., p.160-161):

> *It can be seen that the model predicts slightly better outcomes than have been seen for patients treated with CCR in UK clinical practice. When adjustment is made for subsequent therapies, the survival curves move closer to that seen in the real world. This further emphasises that CCR survival in AZA-AML-001 could have benefited from patients switching treatment to receive azacitidine which they currently cannot do in clinical practice in the UK. The similar curve shapes suggest the model is replicating real life experience plausibly and that the results of AZA-AML-001 can be interpreted with a degree of comfort once adjustments have been made for subsequent treatments.*

Figure 15: Comparison of CCR overall survival predictions to Haematological Malignancy Research Network data

Key: CCR, conventional care regimens; HMRN, Haematological Malignancy Research Network
Source: Celgene submission, Figure 35, p. 160

Figure 16: Comparison of CCR overall survival predictions for patients with poor-risk cytogenetics to Haematological Malignancy Research Network data

Key: CCR, conventional care regimens; HMRN, Haematological Malignancy Research Network
Source: Celgene submission, Figure 36, p. 161

5.3 Critique of company's submitted economic evaluation by the ERG

5.3.1 Critical appraisal checklists

Table 45: NICE reference case

NICE reference case requirement		Meets criteria?	Reviewer comment
Defining the decision problem	The Scope developed by NICE	Y	
Comparator	Therapies routinely used in the NHS, including technologies regarded as current best practice	U	The ERG believe that subsequent treatments after initial treatment with azacitidine, IC and LDAC should have been allowed in the model, which assumed that the no subsequent AML treatment was given and patients were managed by BSC instead.
Perspective on outcomes	All health effects on individuals	Y	
Perspective on costs	NHS and PSS	Y	
Type of economic evaluation	Incremental Cost-utility analysis	Y	
Time horizon	Lifetime	Y	At the mean age, 75, of patients at the start of the effectiveness study and economic model the 10-year horizon used in the analysis covers most patients' remaining lifetime.
Synthesis of evidence on health effects	Based on a single study	U	Based on evidence from a single trial (AZA-AML-001) since systematic review of clinical effectiveness found no relevant studies. The searches for the systematic review were judged to be poorly designed and reported.
Measuring and valuing health effects	QALYs	Y	
Source of data for measurement of HRQL	Reported directly by patients and/or carers	Y	Based on data collected with a disease specific questionnaire (EORTC-Q30) from a single trial (AZA-AML-001)
Source of preference data for valuation of changes in HRQL	Representative sample of the public	Y	EQ-5D survey mapped from disease specific single trial data using a published mapping algorithm

NICE reference case requirement		Meets criteria?	Reviewer comment
Evidence on resource use and costs		Y	Unit costs reflect the perspective adopted. However, resource use was based on experts' opinion on expected quantities of resource utilisation by disease state and initial treatment; it is unclear why those quantities are different across initial arms in the progressive disease state when all patients are assumed to receive best supportive care only.
Discount rate	3.5% p.a. for costs and health effects	Y	
Equity weighting	An additional QALY has the same weight regardless of the other characteristics of the individuals receiving the health benefit	Y	

Key: BSC, best supportive care; HRQL, health-related quality of life; IC, intensive chemotherapy; LDAC, low-dose cytarabine; N, no; p.a., per annum; QALY, quality-adjusted life year; U, unclear; Y, yes

Table 46: Drummond checklist[65]

Item	Critical appraisal	Reviewer comment
Has the correct patient group/population of interest been clearly stated?	Y	
Is the correct comparator used?	N	The ERG believes that subsequent therapy after azacitidine and CCR (IC and LDAC) is likely to be used in the UK, but the model assumed only BSC was given after the initial treatments being compared.
Is the study type reasonable?	Y	
Is the perspective of the analysis clearly stated?	Y	UK NHS PSS
Is the perspective employed appropriate?	Y	NHS Reference Costs
Is the effectiveness of the intervention established?	Y	The trial on which the model is based, AZA-AML-001 establishes the effectiveness of azacitidine relative to CCR in general and the subgroup of people eligible only to BSC in particular.
Has a lifetime horizon been used for analysis, if not has a shorter time horizon been justified?	Y	The model ran for 10 years. Although this is shorter than lifetime given the average patient starting age of 75, most people would be expected to die by the end of the modelled time horizon.
Are the costs and consequences consistent with the perspective employed?	Y	All costs are presented from the UK NHS & PSS perspective
Is differential timing considered?	Y	All future costs and benefits are discounted with a 3.5% rate.
Is incremental analysis performed?	Y	
Is sensitivity analysis undertaken and presented clearly?	N	Probabilistic sensitivity analyses is reported, although it inadequately assesses variations in structural uncertainty due to clinical effectiveness outcomes not adhering to the maintained assumption of the base case relative treatment effects (i.e., proportional hazards).

Key: BSC, best supportive care; CCR, conventional care regimen; HRQL, health-related quality of life; IC, intensive chemotherapy; LDAC, low-dose cytarabine; N, no; Y, yes

Table 47: Philips checklist[66]

Dimension of quality		Critical appraisal	Comments
Structure			
S1	Statement of decision problem/objective	Y	To evaluate the cost-effectiveness of azacitidine in older patients with de-novo or secondary AML with > 30% bone marrow blasts who were not eligible for hematopoietic stem cell transplantation (HSCT), with intermediate- or poor-risk cytogenetics, Eastern Cooperative Oncology Group performance status (ECOG PS) scores 0–2, and white blood cell count ≤ 15×10^9/L
S2	Statement of scope/perspective	Y	NHS & PSS perspective was implemented. Cost and benefit inputs were consistent with this. Scope of the model stated.
S3	Rationale for structure	U	Unclear; see next comment
S4	Structural assumptions	N	Generally, the ERG is not convinced by some of the structural assumptions. These are explored in Section 5 of this report. The model structure is not consistent with routine practice in relation to use of subsequent treatments nor with the adjustment for treatment switching methodology used to derive the hazard ratios that populated the model (i.e., adjustment was only made of CCR arm -and only for subsequent azacitidine treatment-; subsequent treatment use in the azacitidine arm in AZA-AML-001 was not adjusted).
S5	Strategies / comparators	Y	Azacitidine was compared with CCR, which itself was a composite of different preselected treatments according to patient health status and patient and physician preference
S6	Model type	Y	A semi-Markov model. The choice of model type is adequate.
S7	Time horizon	Y	The model ran for 10 years. At the average typical age of this patient population, 75 years, by the end of this period most patients are expected to have died.
S8	Disease states / pathways	Y	Four disease states were modelled two depending on the initial response: Remission (CR/CRi) or Non-remission (PR, SD), Post-Progressive disease/relapse and Death.
S9	Cycle length	Y	Cycle length is 4 months. Clinical opinion sought by the ERG indicated that this cycle length should be appropriate to capture the events and outcomes most influential on costs and quality of life.
Data			

Dimension of quality		Critical appraisal	Comments
D1	Data identification	U	The evidence on relative effectiveness was sourced from AZA-AML-001, which is likely to provide the best available evidence. However, it is possible that other complementary sources of relevant effectiveness evidence may have been missed due to poor quality of methods used for systematically searching bibliographic databases. Searches of costs and cost-effectiveness studies were appropriate.
D2	Pre-model data analysis	U	No information given
D2a	Baseline data	Y	Baseline data used in the model presented similar characteristics to those from the source of effectiveness data, the AZA-AML-001 study, which includes 3.9% of patients with < 30% blasts.
D2b	Treatment effects	N	The ERG feel that the treatment effects may be biased, since a) they are based on overall survival outcomes in the azacitidine arm that are unadjusted for treatment switching, unlike OS outcomes in the comparator arm, and b) the untested and implausible assumption of constant proportional hazards.
D2c	Quality of life weights (utilities)	N	HRQoL was recorded in AZA-AML-001 using a validated disease specific questionnaire (EORTC QLQ30). These outcomes were measured every two cycles; clinical experts consulted by ERG advised that this is unlikely to miss the effects of important acute health events on quality of life. EQ-5D utilities were derived from these disease specific measures using published mapping algorithms. The model grouped together the effect of adverse events and may have thus failed to account for the effects of repeated adverse events. However, this limitation is likely to have limited effect on results.
D3	Data incorporation	N	Parameter values in the submission are well referenced. A number of errors in cell referencing and formulas were found. An error was found in the way parameter values for the number of treatment cycles was inputted in the model, which resulted in a large underestimation in costs of drug administration, monitoring tests, transfusions and the company's drug ICER
D4	Assessment of uncertainty	N	A range of sensitivity analyses was presented, but important sources of uncertainty were left unexplored or unaccounted for.
D4a	Methodological	N	The main analysis (based on IPCW approach) of overall survival outcomes was not informed by specification tests. The analysis of Progression Free Survival was based on the Cox proportional hazards assumption, which was tested and found to be inconsistent but nevertheless used by the company in their base case analysis.
D4b	Structural	N	Structural uncertainty associated with the untested assumption of proportional hazards (PH) in the effect on overall survival was not assessed. Neither the effect of using non-PH to estimate the effect on relapse free survival was investigated.

Dimension of quality		Critical appraisal	Comments
D4c	Heterogeneity	Y	Subgroup analysis was undertaken in the model.
D4d	Parameter	Y	Univariate deterministic and multi-way probabilistic sensitivity analyses were performed.
Consistency			
C1	Internal consistency	N	Even though Celgene claim to have sought validation for the Excel model, the model did not match the number of treatment cycles observed in the single trial that was the source for the relative effectiveness estimates in the model. The model is in any case inconsistent by design with the same source, since the former assumed no subsequent therapy use in the azacitidine treatment group whereas the latter allowed active subsequent treatments in the respective arm.
C2	External consistency	U	Only expert opinion was sought for external validation.

Key: BSC, best supportive care; CCR, conventional care regimen; CR, complete remission; CRi, complete remission with incomplete blood count recovery; HRQL, health-related quality of life; IC, intensive chemotherapy; IPCW, inverse probability of censoring weights method; LDAC, low-dose cytarabine; N, no; PH, proportional hazards; PR, partial remission; PSS, personal social services; SD, stable disease; Y, yes

5.3.2 Model structure

This four-state model structure has the advantage of being simple and transparent. One possible drawback of the model simplicity is that some states are too broadly defined to capture important differences in costs and quality of life between the treatments being compared. For example, it is questionable whether using the same health state to measure costs and health related quality of life for patients who experience relapse or disease progression after initial remission and patients with disease progression after initial non-remission may mask important differences in outcomes between treatments. However, clinical experts consulted by the ERG suggested that combining the Post-Progressive Disease and the Relapsed states into a single health state in the model had clinical face validity.

The main limitation of the model structure was the assumption that no subsequent active treatment was given after the initial azacitidine or CCR treatment. The model assumption that only BSC would be given following initial treatment is questionable since, as advised by clinical expert opinion, patients treated under azacitidine would be likely to receive LDAC (IC as subsequent therapy would not be likely). Similarly, patients under CCR, specifically IC, could be eligible to receive and able to benefit from LDAC.

The absence of all subsequent treatment in the model is also inconsistent with the AZA-AML-001 trial, from which the clinical effectiveness data used in the model was derived. As seen in *Table 48*, a number of subsequent treatments were used in the trial after azacitidine and the CCR treatments.

Table 48: Subsequent therapies used in AZA-AML-001 by randomly allocated therapy

Azacitidine treatment group			CCR treatment group, Best supportive care only			CCR treatment group, Intensive chemotherapy			CCR treatment group, Low dose cytarabine		
Subsequent treatment	**Freq.**	**%**	**Subsequent treatment**	**Freq.**	**%**	**Subsequent treatment**	**Freq.**	**%**	**Subsequent treatment**	**Freq.**	**%**
Azacitidine[a]	9	13.43	Azacitidine	4	66.67	Azacitidine	5	27.78	Azacitidine	22	43.14
Cytarabine	37	55.22	Cytarabine	1	16.67	Cytarabine	9	50.00	Cytarabine	12	23.53
Decitabine	2	2.99	Etoposide	1	16.67	Other[c]	4	22.22	Hydroxycarbamide	6	11.76
Hydroxycarbamide	10	14.93							Mercaptopurine	5	9.80
Mercaptopurine	5	7.46							Other[d]	6	11.76
Other[b]	4	5.96									
Total	67	100.00	Total	6	100.00	Total	18	100.00	Total	51	100.00

Notes: a, The ERG believes this may be a coding error in the company data; b, Includes one instance each of: 'Chemotherapeutics','Erismodegib', 'Gemtuzumab ozogamicin' and 'investigational drug'; c, Includes one instance each of: 'Decitabine', 'Hydroxycarbamide', 'Mercaptopurine', and 'Tioguanine'; d, Includes one instance each of: 'Clofarabine', 'Decitabine', 'Etoposide', Gemtuzumab ozogamicin', 'investigational drug' and 'Tioguanine'

5.3.3 Population

The model considered in the base case a cohort of patients with mean age 75, 59% of whom are male, weighing 71 kg and with a Body Surface Area of 1.80 m². In terms of the pre-selected CCR treatments, the distribution used was that of AZA-AML-001 (BSC, 18%; LDAC, 64%; IC, 18%) but results from sensitivity analyses using the alternative distribution in the HMRN registry (BSC, [redacted]; LDAC, [redacted]; IC, [redacted]) were also submitted for this assessment. The clinical experts consulted by the ERG advised that the distribution of patients amongst CCR treatments in these registry data was different to their local practice but they thought it plausible that the registry data reflected routine UK practice and they had no alternative data.

A footnote to Table 1 of the study report by Dombret et al. 2015,[45] states that "Patients were randomly assigned on the basis of local pathology assessment of baseline BM blast count, which was subsequently reviewed by the central pathologist; in a small number of cases, baseline blast count was <30% upon central review." In the data provided by Celgene, baseline blast count data were grouped in bands, 0 to 5%, 5 to <25%, 25 to <50%, and 50-100%, and 3.9% (19/488) of the sample had baseline blast count <25%.

5.3.4 Interventions and comparators

The specification of evaluated treatments was consistent with randomly allocated treatments specified in AZA-AML-001. The number of cycles of treatment and the doses were intended to mirror those in the trial. However, the way parameter values for the number of treatment cycles were implemented in the model was incorrect, resulting in a mean number of treatment cycles in the azacitidine group of 5.6 instead of the intended 8.8, in the CCR group IC of 1.86 instead of 2 (initiation and consolidation) and 4.4 when estimating drug acquisition costs and 5.3 when calculating the costs of drug administration, tests and transfusion instead of 6.10 in the CCR group LDAC.

In addition, the description of the CCR IC regimen in the economic section of the submission (*Section 5.2.3, p. 72*), states that one-cycle of induction with IC was followed by "two consolidation cycles for those who responded to the treatment, followed by BSC. Those who do not respond to induction therapy receive BSC". In effect, the model applied costs for up to two consolidation cycles to the group of patients who achieved CR/CRi after induction and did not relapse (i.e., were in Remission) and to the group of those who did not achieve CR/CRi after induction but whose disease did not progress (i.e. were in non-Remission).

As discussed before, subsequent therapies were not allowed in the model, despite their use in AZA-AML-001. It is possible that the use of subsequent treatments in the trial may have resulted in a number of treatment cycles that may not correspond with the number of treatment cycles that would be expected in a situation such as that modelled by Celgene, where subsequent therapy is unavailable.

In any case it is questionable that the sequence of treatments studied in the model (i.e., AZA followed by BSC and CCR followed by BSC) is realistic and the relative effectiveness parameter values used in the model themselves reflect a treatment pathway different to that of the model, especially for the azacitidine intervention, whose estimated relative survival effectiveness was not adjusted for the effects of subsequent treatments in the RCT data source as discussed in *Section 5.3.6.*

5.3.5 Perspective, time horizon and discounting

The NHS and PSS perspective was used, in line with the NICE reference case. Given the mean age of the modelled cohort and the limited life expectancy of its patient population, the 10-year time horizon is likely to capture practically all important differences in costs and health benefits as almost all patients would have died within such period. Discounting was also applied to costs and QALYs as in the NICE reference case.

5.3.6 Treatment effectiveness and extrapolation

5.3.6.1 Extrapolation of overall, progression-free and relapse-free survival and effectiveness estimates of progression-free and relapse-free survival

The choice of parametric curves for extrapolation of OS, PFS and RFS, was based on a comparison of goodness of fit statistics associated with the candidate parametric models. However, only models that implied proportional hazards treatment effects were considered (i.e., exponential, Weibull and Gompertz). Other parametric models, in particular log-logistic and log-normal models, i.e., accelerated failure time models, which allow increasing event rates over time at the start of follow-up and decreasing event rates at later times, were not considered.

According to the submission, Celgene's decision only to consider PH models found support in the statistical tests of the assumption for the OS and PFS, but not in the tests results of RFS data. The company states that (Source: Celgene submission, Section 5.3.3, p. 112):

> *Figure 17 [of the company's submission] shows the log-log plot for RFS hazards; unlike OS and PFS, these indicate that the PH assumption is weak, and a Cox regression run with an interaction between treatment group and ln(time) showed a statistically significant effect of the interaction (p-value of 0.011); however, the PH assumption overall has been retained for consistency.*

It is unclear what consistency means in this statement, but it appears the PH assumption was imposed in the extrapolation and estimation of treatment effects on relapse because PH was not rejected in the analysis of OS and PFS, which is a methodology that is likely to be flawed.

The submission elaborates on the company's methodological practice stating that "HRs are also used still for RFS in the model because the shape of the RFS curves, both overall and for treatment groups and subgroups, are not well suited for independent regression models (for illustration of this: there is no indication visually that independent regression models would better characterise observed RFS for extrapolation)". It is unclear what this statement tries to convey. In any case, Celgene had better options to their chosen biased approach. The company could have fitted a separate curve to the RFS data of each trial arm, instead of forcing a common parametric shape on those curves. Also more flexible models could have been estimated than those supporting the proportional hazard assumption, which although convenient was not indispensable in this or any similar analysis.

The analysis of OS in the submission is also flawed. Although the company reported results of a statistical test (for statistically significant interaction of the treatment group variable and the logarithm of time) and visual inspection of log-log plots that supported the PH assumption, these diagnostic checks were only applied to data that was unadjusted for

treatment switching, i.e., had no IPCW weights applied. The company did not perform diagnostic tests on the data underlying the IPCW Cox PH estimates that were ultimately used for the base case economic analysis. We elaborate on this problem below.

Unlike OS and RFS, PFS was the only time-to-event outcome where tests for the proportional hazard assumption supported the PH assumption. These included statistical tests on the interaction of the treatment group indicator and the logarithm of time ($p<0.187$, Source: Celgene submission, Section 5.3.3, p. 114), as well as visual inspection of log-log plots. Celgene also documented the goodness of fit statistics in support of their choice of parametric curve (i.e., Gompertz).

The set of statistics used to inform Celgene's choice of time-to-event curves for the base case analysis (Celgene submission, Section 5.3.4, pp. 120–121), are reproduced below in *Table 49* and *Table 50*. These are goodness of fit statistics for candidate parametric models fitted to each arm for each of the three effectiveness outcomes (OS, PFS and RF). *Table 49* refers to ITT data whereas *Table 50* refers to results of analyses that censor-at-switch to AML therapy data in both arms. The Weibull provides the best, most parsimonious parametric fit to the RFS data (i.e., lowest AIC and BIC statistics). The best fit to the PFS data was the exponential for the azacitidine arm and Gompertz for the CCR, although differences in goodness of fit between models were smaller than those for RFS or OS data. The best parametric fit to the OS data is not that clear for azacitidine, but for the CCR Gompertz appears best for ITT and censor-at-switch data.

Table 49: Goodness of fit for OS, RFS, and PFS parametric functions (ITT data)

Parametric model	OS				RFS				PFS			
	Azacitidine (n = 241)		CCR (n = 247)		Azacitidine (n = 67)		CCR (n = 62)		Azacitidine (n = 112)		CCR (n = 111)	
	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC
Weibull	752	759	799	806	**100**	**104**	**133**	**138**	353	359	373	378
Gompertz	752	759	**793**	**800**	108	113	137	141	353	359	**372**	378
Exponential	**750**	**754**	802	806	149	151	149	151	**351**	**354**	374	**376**

Key: AIC, Akaike information criterion; BIC, Bayesian information criterion; CCR, conventional care regimens; ITT, intention-to-treat; OS, overall survival; PFS, progression-free survival; RFS, relapse-free survival
Note: Figures in bold indicate best fitting model according to criteria referred to by the column heading
Source: Celgene submission, Table 35, p. 120

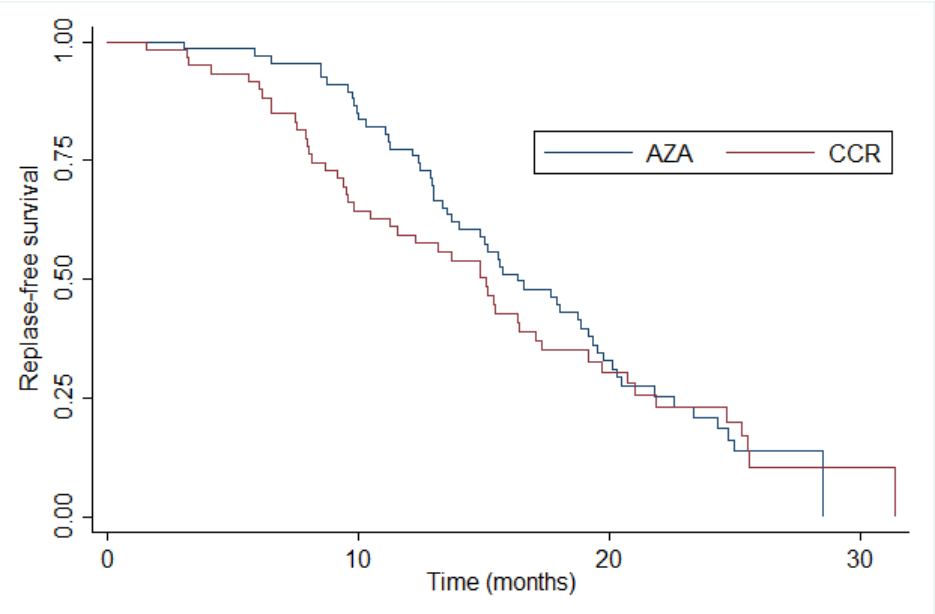
Table 50: Goodness of fit for OS, RFS, and PFS parametric functions (censor-at-switch data)

Parametric model	OS				RFS				PFS			
	Azacitidine (n = 241)		CCR (n = 247)		Azacitidine (n = 67)		CCR (n = 62)		Azacitidine (n = 112)		CCR (n = 111)	
	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC
Weibull	**674**	681	694	701	**103**	**108**	**137**	**141**	344	349	344	349
Gompertz	674	681	**684**	**691**	113	117	141	145	342	348	**342**	**347**
Exponential	676	**680**	700	704	150	152	150	152	**342**	**345**	344	347

Key: AIC, Akaike information criterion; BIC, Bayesian information criterion; CCR, conventional care regimens; OS, overall survival; PFS, progression-free survival; RFS, relapse-free survival
Note: Figures in bold indicate best fitting model according to criteria referred to by the column heading
Source: Celgene submission, Table 36, p. 121

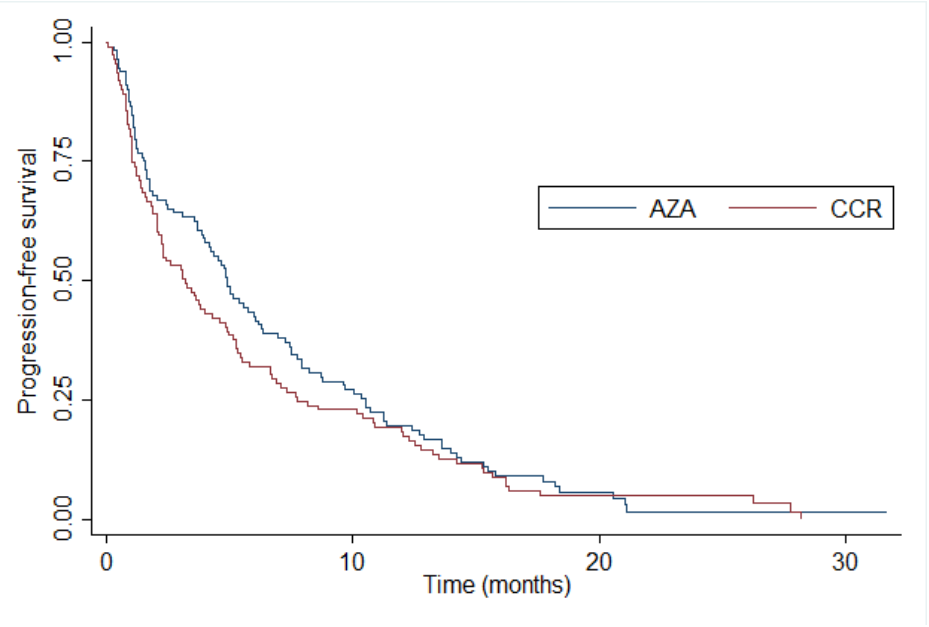
The range of models evaluated by Celgene was too restrictive, as these are all models with constant or monotonic hazards that are assumed to be proportional between the two treatments. Since, as discussed above and in the submission, RFS failed the proportional hazards test, fitting separate Weibull curves, as the company did for its base case analysis and ERG verified in the Excel model files, is still likely to bias the extrapolation of RFS outcomes. In fact, seeking to fit a rigid statistical model to limited RFS data (n=67 in azacitidine and n=62 CCR at randomisation) and PFS (n=112, and n=111, respectively) lacks justification because during the observed trial period most people relapse and progressed (*Figure 17* and *Figure 18* respectively).

Figure 17: Kaplan-Meier estimate of relapse-free survival in AZA-AML-001



Key: AZA, azacitidine; CCR, conventional care regimens
Source: Celgene submission, Figure 16, p. 113

Figure 18: Kaplan-Meier estimate of progression-free survival in AZA-AML-001

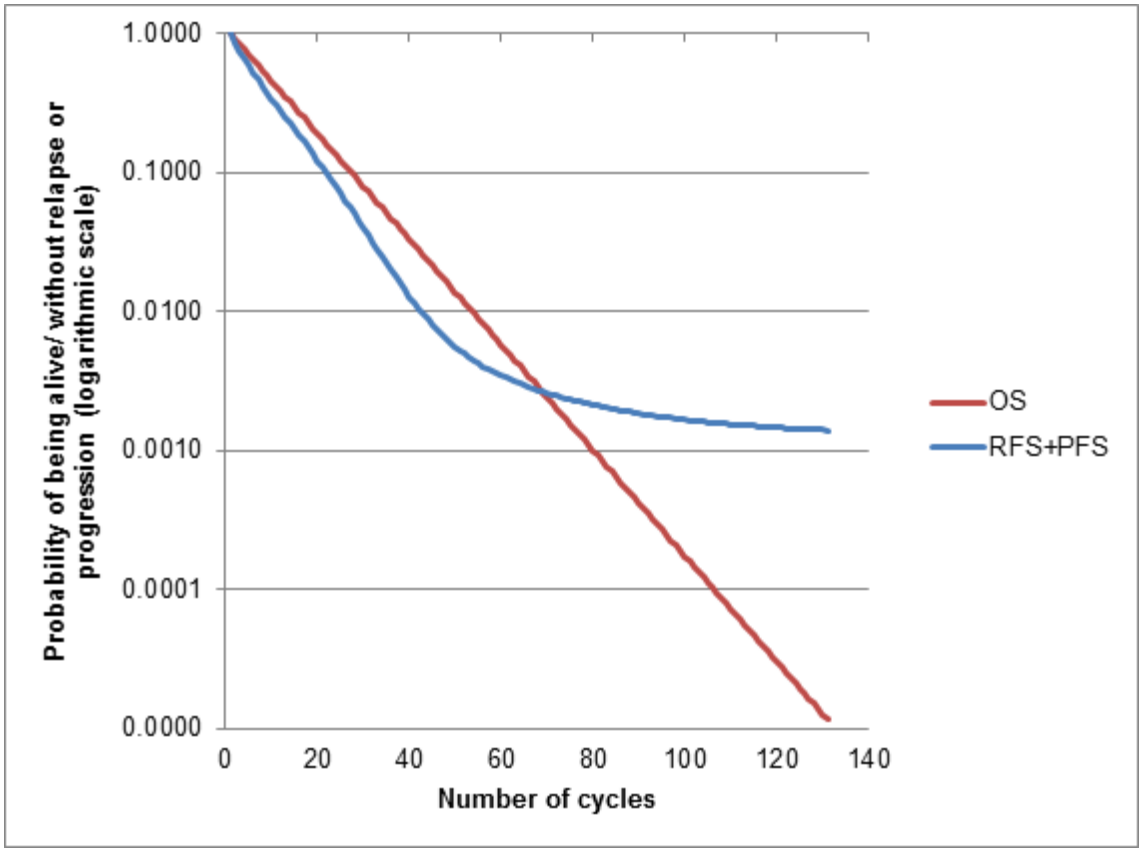


Key: AZA, azacitidine
Source: Celgene submission, Figure 18, p. 115

The inadequacy of the statistical analysis may be illustrated by comparing the estimated time to event curves used by Celgene in their base case analysis, combining PFS with RFS and contrasting it with OS. *Figure 19* below illustrates the problem for the CCR arm. By cycle 69, that is, just after the start of the sixth year after first receiving treatment the estimated curves imply that there are more patients in the treated cohort who are either in Remission or non-remission (stable disease) than there are patients alive. Ironically, Celgene's consistent choice of parametric extrapolation curves that implied PH (possibly including their choice of overall survival curve) led them to use survival models that are mutually incompatible.

It is tempting to disregard the problem depicted in *Figure 19* by thinking that the crossing of curves occurs only when less than 1% of patients are still alive. However, the crossing itself is a sign that the estimated time spent by patients in the different states of the model (and the associated base case results) may be severely biased. In fact, it is easy to see why the two curves in *Figure 19* have incompatible shapes: the OS curve assumes a constant hazard (which appears as a straight line in the figure due to log scale metric used for the Y axis), whilst the RFS was modelled using a Weibull function with a shape parameter estimate of 1.7 (Celgene Excel model sheet 'CCR parameters' cell HF42), which implies an increasing hazard; the shape of these two curves would be mutually compatible as time passes, but not accompanied by the Gompertz function used for PFS since its estimated negative parameter of -0.03 (Celgene Excel model 'CCR parameters' cell HF30) implies a (small) proportion of people never die, which is what ultimately drives the crossing of the curves.

Figure 19: Survival curves used in Celgene's base case (Y-axis in logarithmic scale)



Key: OS, overall survival; PFS, progression-free survival; RFS, relapse-free survival
Source: Produced by the ERG using Celgene's decision model

As a consequence the ERG believes that using Kaplan-Meier nonparametric curves as observed in the AZA-AML-001 trial provide the best source data with which to populate PFS and RFS model parameters, while minimising the structural uncertainty of the cost-effectiveness results.

As for OS, it is not clear to the ERG whether the data underlying these assessments of parametric curves are censored at switch to some or all subsequent treatment use, and to CCR only or both arms. Analyses based on censoring for switching in both arms, possibly adjusting for non-random treatment switching may be required to obtain valid estimates of survival benefits, and the choice of parametric extrapolation need not restrict to proportional hazards functions without statistically testing for such assumption in the data. Therefore, the ERG undertook further analyses of individual patient data provided by the company; these analyses are presented in the next section, where the justification for extrapolating using parametric curves is also considered.

5.3.6.2 Critique of adjustment of overall survival for subsequent treatments

This section is focused on the IPCW method used by Celgene to derive its primary estimates of relative effectiveness. The other methods explored in the submission either faced problems of face-validity (the RPSFTM and IPE methods), were not feasible (the two-stage method, which required a second baseline), or appeared only as an appendix (Celgene Post hoc statistical methods addendum AZA-AML-001) without details of their application to the present assessment or results.

As discussed in *Section 5.2.5.1* (*page 77*), the validity of the IPCW method, which originated in the literature of causal effects estimation using observational data, is limited by its requirement that data for all relevant confounders related to treatment switching and mortality are included in the analysis, and tends to perform poorly in small samples or applications with rare events. In an RCT context where the method could take advantage of high quality prognostic information to adjust for observed confounding in treatment switching, the IPCW method is hampered by the small size of available samples of patient data. Consequently, IPCW was not used by Celgene to estimate OS effectiveness by CCR preselected therapy.

Indeed, the company stated (Source: Celgene submission, Section 5.3.5, p. 123):

> *subgroup adjustment was not feasible because of limited data on switching; however, a clinical expert consulted during this analysis stated that questions can be raised about the clinical generalizability of the results in subgroups [This refers to controlling/adjusting for the three treatment groups within CCR, as well as for cytogenetic risk or MDS subgroups, as clarified by Celgene in response to questions by ERG], because clinicians can identify potential switching candidates based on observed performance, and recommended focusing on the adjusted data for overall patients.*

Similarly, in response to the request by ERG to clarify the reason for not adjusting event-free survival, relapse-free survival and progression-free survival for treatment switching, the company answered that "This was due to sample size primarily. The instances in which switching preceded the clinical event of interest were few, and the impact of this on the results would be very small."

Celgene provided the ERG with the individual patient data for replicating the IPCW estimation of the Cox proportional hazard ratios used in the company's base case cost-effectiveness analysis. ERG was able to replicate those results (*Table 51*). The ERG was unable, however, to replicate the estimation of IPCW weights, which were provided by the company but without the dataset used to estimate them. Thus the ERG can only comment on the quality of the analyses of OS with the IPCW weights as given, without being able to assess whether the statistical model used to estimate these weights is of good quality. This is a relevant issue given the relatively small sample available for analysis and the large number of groups of patients as defined by the covariates used in the IPCW model, which according to the SAS code (Celgene appendices to submission, Appendix 10) and *Table 52*, included four binary and one three-level fixed baseline variables, three binary time-varying covariates, and one variable indicating the 76 15-day periods of observation, for a total of $2^4 \times 3^1 \times 2^2 = 192$ possible subgroups at each of the 76 follow-up times. It is unlikely that the validating condition (see *Section 4.3.2, p. 61*) of there being a positive probability of not being censored at each and every follow-up point and for every combination of values observed for the covariates in the IPCW model, might have been met in this sample, and this became less likely as time passed in the trial.

Table 51: Results of the IPCW models

Model	HR (95% CI)	P value[a]
Unadjusted	[redacted]	[redacted]
Adjusted	[redacted]	[redacted]

Key: HR, hazard ratio; IPCW, inverse probability of censoring weight
Notes: a, P value calculated using a log-rank test
Source: Celgene submission, Table 38, p. 126

Table 52: List of covariates used for calculating stabilised weights in the IPCW model

Non time-varying covariates	Time-varying covariates
• [redacted]	• [redacted]

Key: AML, acute myeloid leukaemia; CR, complete remission; CRi, complete remission with incomplete blood count recovery; ECOG, Eastern Cooperative Oncology Group
Source: Celgene appendices to submission, Appendix 11, Table 14

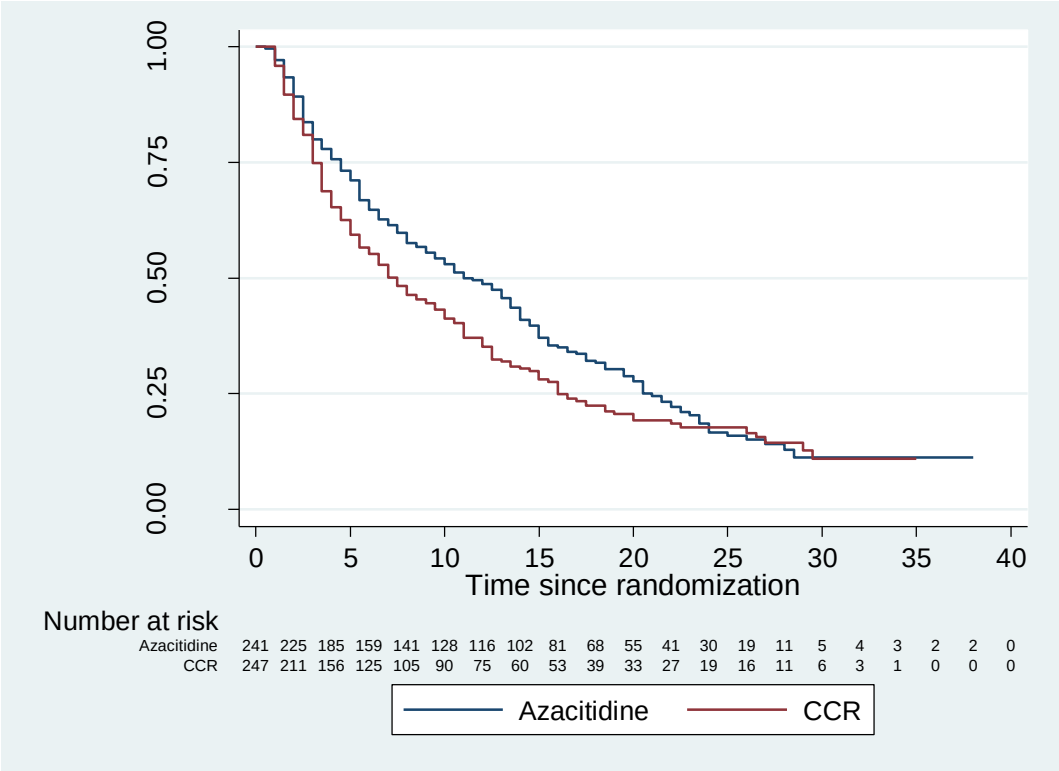
The IPCW estimate of OS effect, which was selected for the company's base case cost-effectiveness analysis, is based on the assumption of proportional hazards. Celgene acknowledge this limitation (Source: Celgene Submission, Section 5.3.5., p. 122):

> *A further limitation of the IPCW approach is that it does not produce counterfactual survival data directly; in order to use adjusted CCR data in subsequent survival analysis and economic modelling, counterfactual data are required. The survivor function approach, in which the CCR hazard function is calculated using the observed azacitidine hazard function and using the inverse of the IPCW-adjusted HR, changed the shape of the CCR survival curve relative to the observed data. On the face of it, this is problematic – as well as altering the shape of the CCR survival curve, the method forces hazards to be proportional – but it must be acknowledged that the purpose of the analyses is to produce counterfactual data, which may result in counterfactual hazard functions (i.e., different KM curve shapes) and not just counterfactual hazards from curves whose shapes do not change.*

It is not clear, however, why Celgene did not test for the proportional hazard assumption using the IPCW adjusted data. In their OS analysis, they overlooked the possibility of graphically and statistically testing for the constant proportional hazards assumption. They seemed to ignore the fact that the survival curve of the CCR arm under the IPCW adjustment is observed; in *Figure 20*, we plot the CCR OS Kaplan-Meier curve for the CCR arm after adjustment for IPCW. In contrast, *Figure 21* depicts the curve used by Celgene in

their base case analysis. The Kaplan-Meier curve for the azacitidine arm is one and the same since it was not subject to adjustment for treatment switching by the company. This suggests that forcing the proportional hazards assumption on the OS curves is likely to result in different estimates of survival benefits from those obtained by fitting separate parametric curves, of possibly different shapes, to the OS data from each of the two trial arms. For the base case, Celgene used a hazard ratio estimated from a Cox proportional hazards analysis adjusted for baseline covariates. The ERG undertook diagnostic tests of the proportional hazards assumption on the adjusted Cox PH OS curves by comparing the cumulative log-log plot of the azacitidine and CCR arms directly as well as testing for time and treatment effect interactions and Schoenfeld residuals (see details in *Appendix 2*). The results of these tests suggest the constant proportional hazards assumption is not supported by the data.

Figure 20: Overall survival curves in AZA-AML-001 following adjustment of the CCR arm for subsequent treatment with azacitidine using the IPCW method



Key: CCR, conventional care regimens; IPCW, inverse probability of censoring weight; ITT, intention-to-treat
Note: Based on discrete (15-day period) time-to-event data
Source: Produced by the ERG using individual patient data provided by Celgene

Figure 21: Overall survival from AZA-AML-001 adjusted using hazard ratio from the IPCW method

Key: CCR, conventional care regimens; HR, hazard ratio; IPCW, inverse probability of censoring weight
Source: Celgene submission, Figure 25, p. 127

Celgene stated that progression and relapse events in the trial occurred before other censoring events and few cases of progression and relapse occurred after subsequent therapy use in CCR (Source: Celgene responses to NICE questions for clarification). This suggests that adjustment of PFS and RFS using the IPCW method would not have been practicable due to very small numbers, but the company could have applied simple censoring at switch adjustments on curves to verify that indeed subsequent therapy use was not important for the results, as the company suggests.

In any case, these adjustments for treatment switching are inconsistent with the economic model for which they were intended. The Celgene model assumes no subsequent therapy use was available after either azacitidine or CCR (Celgene submission, Section 5.6.2, p. 161). In contrast, the clinical effectiveness analyses conducted by Celgene only adjusted for subsequent treatment with azacitidine in CCR arm, thus ignoring the use of subsequent treatments in the azacitidine arm of AZA-AML-001 (the source of the clinical effectiveness data), as well as any active subsequent treatments other than azacitidine used in the CCR arm.

In defence of their methodology, Celgene referred to the methodological recommendations in the NICE DSU TSD16[51] as indicating that the NICE preferred approach is to adjust treatment switching in the comparator (CCR arm) but not the intervention (azacitidine arm). As discussed in *Section 4.3* (*page 59*), this is an incorrect interpretation of NICE DSU TSD16, which clearly recommends that in situations where, as in the present case, the decision problem involves the evaluation of two alternative states of the world where no subsequent active treatment is available, then adjustment of outcomes for active treatment switching in both trial arms is warranted. Whether the assumption that no subsequent active treatment would be available to UK patients in routine practice is plausible may of course be questioned, but the point here is that applying IPCW adjustment for subsequent treatment to

the outcome data of the CCR but not the azacitidine arm in the AZA-AML-001 trial is inconsistent with the economic model which the overall survival IPCW analysis was designed to inform.

To remedy the contradiction between the model structure, on the one hand, and the methodology underlying the OS treatment effect estimates, on the other, two options were available to the ERG. One was to correct the model to include the costs of subsequent treatments used in the AZA-AML-001 and left unadjusted for in the statistical analysis that produced the base case OS treatment effect estimates. This option was not feasible because the required data on the dates of start and end of subsequent treatments as well as treatment dosages and frequencies of administration were not available to ERG. The alternative of adjusting the estimates of relative effectiveness for subsequent AML treatment use in both arms of AZA-AML-001 was feasible, since the company provided results of such analysis using IPCW (Celgene submission, Table 21, p. 70). A limitation of this option was that statistical tests to validate these analyses were not reported, and the Kaplan-Meier or individual patient data required to perform them were not available to ERG either (not all the required data were available in the individual patient data that Celgene provided for ERG use). Ideally, testing for the Cox PH assumption would be performed and, if rejected, treatment effects estimated using a different model.

The only option available to ERG was to base the test of the proportional hazards assumption on OS data censored-at-switch, which had similar Kaplan-Meier plots as the OS IPCW weighted data in the CCR arm, and resulted in small differences in hazard ratio estimates (Celgene appendices to submission, Appendix 11). The data provided by Celgene did not allow ERG to extend the IPCW analysis to adjust for subsequent AML therapy to both treatment groups. However, the company provided the data to perform censor-at-switch analysis for any subsequent AML therapy use in both trial arms, as well as testing for the proportional hazards assumption in the data. Fitting a range of parametric curves including proportional (exponential, Weibull, Gompertz, Bathtub) and non-proportional hazard parametric models (log-normal and log-logistic) to OS data from each trial arm separately resulted in the following goodness of fit statistics and test statistics for nested models (i.e., whether the simple exponential model could be supported by the results of more complex proportional hazards models).

Table 53: Goodness of fit and test statistics for OS parametric functions (patients censored at switch for any subsequent therapy in both arms)

Parametric model	Unadjusted for baseline covariates						Adjusted for baseline covariates[a]					
	Azacitidine (n=241)			CCR (n=247)			Azacitidine (n=241)			CCR (n=247)		
	AIC	BIC	p^b	AIC	BIC	p^b	AIC	BIC	p^b	AIC	BIC	p^b
Weibull	**676**	687	0.050	696	707	0.008	638	708	0.858	642	713	0.708
Gompertz	**676**	687	0.057	686	697	<0.001	638	708	0.810	641	711	0.202
Exponential	678	**685**		702	709		**636**	**702**		**641**	**707**	
Bathtub	678	689		706	716		640			645		
Log-logistic	676	686		684	694		637	707		643	714	
Log-normal	678	688		**678**	**689**		641	711		644	714	

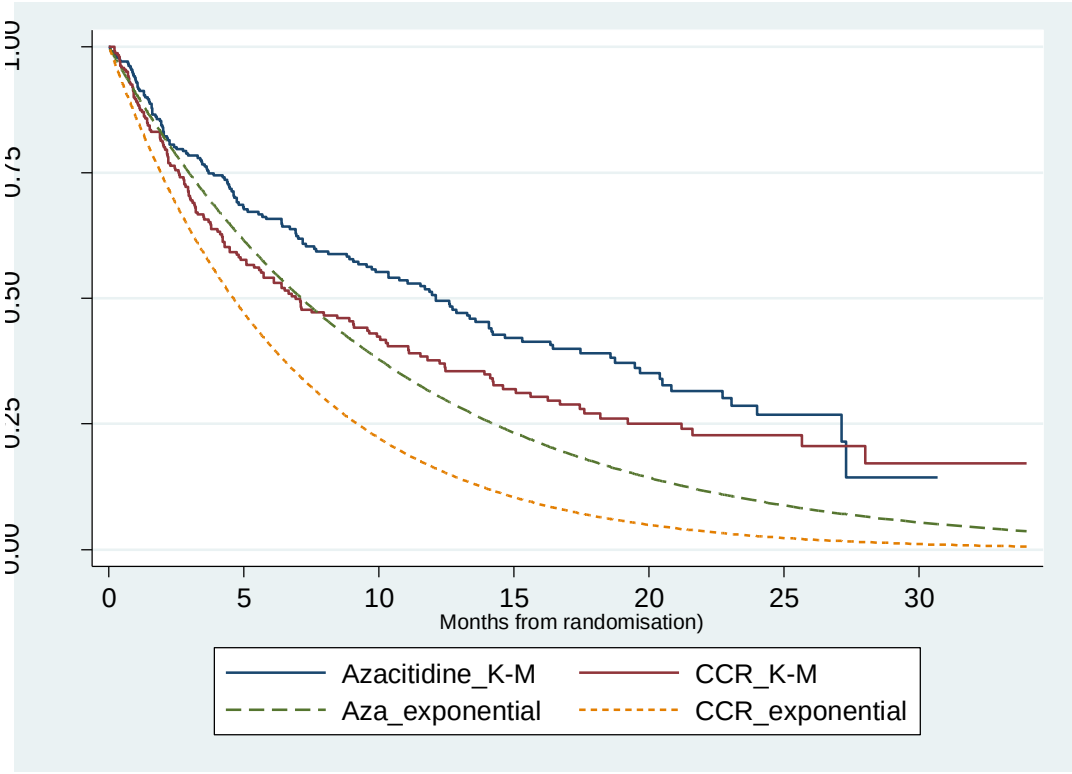
Key: AIC, Akaike information criterion; AML, acute myeloid leukaemia; CCR, conventional care regimens; ECOG, Eastern Cooperative Oncology Group; OS, overall survival

Notes: a, Covariates included in adjustment: age, sex, ECOG, cytogenetic risk, % blast group, CCR therapy preselect group, comorbidities, AML days, platelet transfusion status, geographical region; b, p-value from Z-test on shape coefficient (versus null hypothesis of exponential distribution); Figures in bold indicate best fitting model according to criteria referred to by the column heading

According to AIC, there is no difference in performance between unadjusted models of azacitidine, since all lie within a 2-unit difference of one another.[67] For unadjusted analyses of CCR, the models consistent with PH are not supported by the data, except for the Gompertz, because they perform worse than the model with the minimum AIC, the log normal, by more than 10 units.[67] BIC rewards parsimonious models relative to AIC, since the former penalises the inclusion of each additional covariate in the model by 5.5 points whereas the latter does it by 2. Thus, the exponential, which uses one fewer parameter than the other models, moves up the BIC performance ranking to be the best unadjusted model fit to azacitidine OS data, but still 'strongly' underperforms the unadjusted model that best fits the CCR data, the log-normal model, since it has more than 10 additional points.[68] The best performing unadjusted model for the CCR arm that is consistent with the PH assumption is the Gompertz, but its 8 additional points over the log normal model may be considered 'strong' evidence against it and in favour of the log-normal model.[68] Although the accelerated failure time models (the log normal and log-logistic) perform better than other models they imply implausible predictions as detailed in *Appendix 2*. In contrast, an exponential unadjusted model fitted separately to each arm produces life expectancy estimates of 17.09 versus 12.36 months for azacitidine and CCR, respectively.

As for models that adjusted for baseline covariates, the exponential is the optimal model for both azacitidine and CCR arms, although all other models except the log-normal show comparable AIC and BIC performance. The adjusted exponential model fitted to both arms results in a HR of 0.64 and has a predicted difference in OS of 3.64 months, in favour of azacitidine; details are presented in the *Appendix 2*. *Figure 22* presents the ITT Kaplan-Meier data and the fitted adjusted exponential OS model to data censored at switch to subsequent AML therapy in both trial arms.

Figure 22: Overall survival in AZA-AML-001 – ITT Kaplan-Meier data and adjusted exponential model fitted to censor-at-switch (any AML therapy) data



Key: AML, acute myeloid leukaemia; Aza, azacitidine; CCR, conventional care regimens; ECOG, Eastern Cooperative Oncology Group; ITT, intention-to-treat; K-M, Kaplan-Meier

Note: Adjusting covariates: age, sex, ECOG, cytogenetic risk, CCR therapy preselect group, comorbidities, AML days, platelet transfusion status, geographical region were effects-coded so that baseline is an estimated overall mean in the sample; % blast group <25% adjusted using 1 vs. 0 indicator (see *Appendix 2*)

Source: ERG analysis using censored-at-switch to subsequent AML therapy (in both trial arms) individual patient AZA-AML-001 data provided by Celgene

Although the frequency of subsequent therapy use was similar in both groups, overall (azacitidine 28% versus CCR 30%), and by subgroups (see *Table 54*, Source: Celgene's response to questions for clarification), patients who used subsequent treatments did so earlier in CCR than azacitidine (273 versus 322 days post-randomisation; Source: ERG extraction from individual patient data provided by Celgene). The [redacted] difference between unadjusted censor-at switch (HR 0.72, Source: ERG analysis using individual patient data provided by Celgene) and unadjusted IPCW (HR [redacted], Celgene submission, p. 70, Table 21) for subsequent AML treatment in both arms, suggests the censor-at-switch estimate is biased in favour of azacitidine, and thus may be considered a conservative choice of estimates given the effect on results of other changes adopted by the preferred ERG base case analysis. *Figure 22* depicts the adjusted fitted exponential survival curves alongside the ITT Kaplan-Meier curves.

Table 54: Number of patients who switched treatments in AZA-AML-001

Patient subgroup	Number of patients who did not receive subsequent AML therapy	Number of patients who received subsequent AML therapy
All patients	[redacted]	[redacted]
	[redacted]	[redacted]
Comparator: IC	[redacted]	[redacted]
	[redacted]	[redacted]
Comparator: LDAC	[redacted]	[redacted]
	[redacted]	[redacted]
Comparator: BSC	[redacted]	[redacted]
	[redacted]	[redacted]
Subgroup: Intermediate cytogenetic risk	[redacted]	[redacted]
	[redacted]	[redacted]
Subgroup: Poor cytogenetic risk	[redacted]	[redacted]
	[redacted]	[redacted]
Subgroup: With MDS-related changes	[redacted]	[redacted]
	[redacted]	[redacted]
Subgroup: Without MDS-related changes	[redacted]	[redacted]
	[redacted]	[redacted]

Key: AML, acute myeloid leukaemia; AZA, azacitidine; BSC, best supportive care; CCR, conventional care regimen; IC, intensive chemotherapy; LDAC, low-dose cytarabine; MDS, myelodysplastic syndrome

The ERG believe that baseline covariate-adjusted censor-at-switch methods coupled with parametric curves fitted to each arm offer a practical, transparent alternative to IPCW with similar performance in this application. While the censor-at-switch assumption that subsequent treatment occurs at random is likely biased, in the current case the alternative of adjusting for non-random treatment switching with the IPCW was not feasible due to the the ERG lacking the data required to assess or replicate the model used to estimate the IPCW weights. In any case, given the relative magnitudes of costs and QALYs presented in *Section 6* (*page 130*), the degree of uncertainty in effectiveness parameters is unlikely to matter for the economic results.

As a conservative approach, the ERG's preferred base case analysis adopted the exponential OS HR estimates adjusted for baseline covariates used by the IPCW method, including sex, age, ECOG status, preselected CCR treatment, time since initial AML diagnosis, comorbidity score, in data censored at switch to any subsequent AML treatment from both trial arms in the dataset provided by Celgene to ERG. The resulting HR estimate, 0.64, is more favourable to azacitidine than the respective estimates from applying IPCW to subsequent azacitidine use in CCR only, [redacted] and [redacted] (Celgene's base case), IPCW to any AML in both arms arms, 0.77 and 0.71, and the ITT value of 0.85.[45]

It must be noted that the adjusted censored at switch analysis performed by the ERG in obtaining its preferred estimate of OS effectiveness was not extended to subgroup analyses, since the small sample is likely to lead to bias due to overfitting when adjusting for baseline covariates. The subgroup analyses were thus based on the original censor-at switch exponential OS curves estimates used by the company in their model.

5.3.7 Health-related quality of life

A number of problems were found with the assessment of health related quality of life in the economic analysis. However, based on univariate sensitivity analyses presented by Celgene and conducted by ERG, the bias originating from this element of the model has an insignificant influence on the ICER.

In line with RCT studies in this field, health related quality of life was measured at the start of every other cycle in AZA-AML-001 using a disease specific HRQoL measure, the EORTC QLQ-C30. Since EQ-5D outcomes were not measured in the trial, peer-reviewed mapping algorithms were identified and used by Celgene in accordance with NICE methodological guidance.[41]

Three potential limitations of this analysis were identified by ERG. One related to the possibility that because HRQoL was measured every 56 days in AZA-AML-001, the effects of some important acute adverse clinical events, especially treatment-related ones, may have been missed by the data collection. Advice sought from clinical experts suggests that no obvious important events would have been expected to be missed by the quality of life study supporting the Celgene submission.

Another potential issue was missing data. Celgene reported that longitudinal statistical analysis using mixed methods were employed to assess HRQoL (Celgene Submission, Section 4.7.5.8, p. 77), and in response to clarification questions by ERG, the company stated that:

> *A mixed effect model repeat measurement (MMRM) model was developed for the EORTC QLQ-C30 Fatigue domain with the inclusion of a fixed-effect covariate indicating whether a transfusion had taken place up to 5 days before health related quality of life (HRQL) assessment. This analysis was undertaken because blood transfusions are likely to affect fatigue, but this relationship would not have been explored by previous analyses.*
>
> *Additional MMRM models were developed for the secondary HRQL domains (Dyspnoea, Physical Functioning and Global Health Status/ quality of life (QoL)) that included RBC or platelet transfusion up to 5 days before the HRQL assessment as a factor.*
>
> *The statement 'A mixed model analysis failed to reveal any statistically significant differences in the impact of treatment on all domains between treatment arms' references the post-hoc MMRM analysis controlling for the impact of red blood cell (RBC) or platelet transfusion received up to 5 days before HRQoL assessment. It was hypothesized by the clinical study team that transfusions administered shortly before HRQoL assessment may have an effect on fatigue, and this effect would not have been captured in the initial model. The results of the MMRM analysis for the Fatigue domain without this additional covariate were significant in favour of CCR.*

> *No significant differences were observed for the secondary domains. Full presentation of both results can be located in the CSR section 11.4.1.2.10.7.*
>
> *All MMRM analyses were based on the assumption that data are missing at random. A post-hoc sensitivity analysis utilizing a pattern-mixture model was conducted to explore the impact of the missing-at-random assumption. Results of this analysis aligned with the MMRM results for Physical Functioning, Dyspnoea and Global Health Status/QoL, with no differences between treatment groups at p<0.05, while results favoured CCR for the Fatigue domain (P=0.025).*

The ERG is satisfied that the statistical analyses of HRQoL conducted by Celgene were thorough and conducted following best methodological practice. As is typical with patient reported outcome studies conducted alongside RCT in this clinical area, missing data is a problem due to the high drop-out rates and item non-response, and analytical methods employed by Celgene have tried to address it. However, the results of this analysis were crudely applied in the cost-effectiveness model, as explained next.

The third limitation was the way the effect of adverse events was implemented in the model, which treated disutility effects of adverse events differently from their cost impacts. The ERG asked the company to explain why the adverse events which are costed on page 130 of the company's submission appear to differ from the adverse events for which disutilities are measured (page 129 of the company's submission). The company replied:

> *HRQL analysis from the trial was more restricted in terms of measuring and mapping from EORTC QLQ-C30 scores during an AE in the trial; costing on the other hand used rates of AEs, disaggregated by type, from the main clinical study report, and hence were more detailed.*
>
> *Thus, in the model, AE disutilities are a single figure that are aggregated at the trial-analysis level; AE costs on the other hand are aggregated within the model itself, calculated from rates and unit costs for AEs.*

The QALY impact of AEs was modelled as the probability of at least one TEAE of Grade > 3 per 100 person-years multiplied by the utility of grade > 3 TEAEs. In the model, the effect of AEs on QALYs in azacitidine have been modelled for a maximum of 8.8 treatment cycles, which is effectively less than the mean 8.8 in the trial, while for CCR as a whole they were counted for a maximum number of cycles of 5.1 which is more than the mean number of 2 cycles with IC. These calculations are likely to overestimate the additional costs of CCR relative to azacitidine but this bias had a small effect on the results given other issues identified by the ERG in the model.

These features of the model are likely to lead to bias in estimating the costs and QALYs differences between the azacitidine and CCR treatment groups, as they conflated different types of adverse events and inaccurately measured repeated or continued episodes of AEs.

5.3.8 Resources and costs

5.3.8.1 Drug costs

The ERG identified some errors in the calculation of drug costs. The most important error was that the model incorrectly applied the mean number of cycles reported in the trial.[45] In order to obtain the mean 8.8 number of azacitidine cycles, a maximum number of 19 cycles needs to be set given the distribution of the modelled cohort between the Remission and Non-remission states (the proportion of cohort members in the relapse/PD state and Death states do not consume azacitidine medication). Instead Celgene applied a maximum of 8.8 (rounded down to the closest integer, 8) number of treatment cycles effectively accounting for a mean number of azacitidine cycles in the model less than 6 as opposed to the intended 8.8. Likewise the correct maximum number of cycles for LDAC in CCR is 10 cycles (the 1.6 cycles used for IC in CCR modelled by Celgene is practically correct). In correcting the probabilistic sensitivity analysis ERG chose SE 3 for azacitidine, 1 LDAC and (0.06 for IC as used by Celgene's base case is correct), to calibrate the respective model outputs to the reported figures for mean minus one standard error in the number of cycles reported by Dombret et al. 2015.[45] As documented below, correcting this error of implementation to calibrate the model outputs with the summary statistic reported by the trial increases the ICER from a level around £20,000 to £84,000 per QALY gained.

A mistake was also found in the calculation of dosing. Celgene incorrectly estimated daily dose of cytarabine ('HRU_costs'!C115:C117) assuming mg/m² instead of using mg/day as in the CSR. However, this error does not affect acquisition costs, since the inflated dose is still less than the size of a vial, but has a small effect on ICERs for the 'no wastage' analyses.

More importantly, the costs of drug acquisition and administration, monitoring tests, and transfusions in CCR were based on a formula with reference to the azacitidine number of cycles (8.8) instead of the number of cycles of CCR treatments (2 cycles, 1 initiation and 1 consolidation cycle in IC, and 6.1 in LDAC).

The ERG also corrected an error in how the model calculated the costs of drug administration, monitoring tests and transfusions in the first (induction/pre-response) cycle for both azacitidine (Model AZA sheet AB23:AD23 and Model CCR sheet AB23:AD23). In the first cycle, the model accounts for costs of two cycles for these costs, unlike for drug acquisition costs (AA23 in Model CCR and Model AZA sheets), which only accounts for the initial cycle. This has the effect of loading the costs of two cycles of treatment to all patients, since the survival curves in the model assume all are alive and under treatment while awaiting the initial evaluation of response. For IC this implies one induction cycle and one consolidating cycle thus overestimating the mean number of consolidation cycles reported in the trial of 1.[45]

5.3.8.2 Health resource utilisation and unit costs

The quantities of resource use for medical staff costs, drug monitoring tests and outpatient procedures (including transfusions), and inpatient hospitalisations by health state and initial treatment (azacitidine, IC, LDAC or BSC) were derived from expert opinion. Apart from drug monitoring testing and outpatient procedures during Remission and Non-remission phases (i.e., while the initial treatments are being administered or when treatment has recently been withdrawn or concluded), it is not clear the rationale for having different resource use quantities within the same health state for the different treatments being compared. Models

of health state transitions that use generic health states, where health status entirely determines health related quality of life and costs and where treatment is irrelevant (apart from explicit treatment-related events), are more transparent and arguably robust to bias. In terms of resource use none of the health states are generic, and only adverse events-related costs are independent of initial treatment.

Without a clear clinical rationale, asking expert opinion about costs in the Relapse/progressive disease phase contingent on initial treatment, despite the model assumption that all patients in this phase are managed under BSC only, is susceptible to framing bias in surveyed responses. This issue is apparent in the health resource use questionnaire used by Celgene and presented in its submission Appendix 12. Furthermore, in this model, variation of costs in Relapse/progressive disease across initial treatments are not accompanied by a corresponding variation in utilities, which implies resources are being used without noticeable effects on quality of life.

Most notably in the differences in resource use quantities is the amount of inpatient days per 4-week cycle of 1.73 in azacitidine versus 2.61 in CCR (Celgene Submission, Table 46, p. 134). At the Intensive Care per inpatient day cost used by Celgene of £714, inpatient costs differences between groups accumulate at a rate of £628 per month per patient in Relapse/PD. The CCR per cycle inpatient days figure breakdown by treatment is, as presented in the Excel model file, 1.66, 0.95 and 0.00, for IC, LDAC and BSC (sheet 'Default values' cells I283, M283, Q283). While heterogeneity by pre-selected CCR therapy is to be expected it is unclear why the weighted average across subgroups should differ between azacitidine and CCR in Relapse/PD managed with BSC.

The ERG believes that costs in the Relapse/PD phase should be equal across treatment arms and pre-selected therapy arms, which significantly increases the ICER for azacitidine versus CCR. *Appendix 3* presents the costs per cycle in PD/Relapse under BSC, which were applied to the PD/Relapse state across all arms and therapies within CCR to reflect the assumption that all patients are in the same health state and managed equally.

Another limitation was that the costs of managing AEs in patients from azacitidine and CCR groups were estimated as the product of the probability of at least one TEAE of Grade > 3 per 100 person-years and the average cost of managing grade 3 or 4 TEAEs observed in the AZA-AML-001 trial. The average cost was calculated as arithmetic average of the treatment costs of anaemia, neutropenia, febrile neutropenia, thrombocytopenia, pneumonia and worsening AML. Furthermore, it was assumed that the most frequent TEAEs of grade 3 or 4 and their incidence rates in azacitidine and CCR groups are the same.

In the executable model, the costs of managing AEs in azacitidine and CCR patients have only been accounted as a one-off initial cost, which is at odds with the way the associated quality of life effects were implemented in the model (see *Section 5.3.7*).

5.3.9 Cost-effectiveness results

The deterministic base case ICER presented by Celgene is £20,648 per QALY. Comparison of Tables 56 on QALYs and 57 on Costs results in the submission, (Celgene Submission, Tables 56 and 57, p. 146-147), reproduced below, points to RFS and Relapse/PD as the phases in the model where the largest outcome differences between treatment groups appear. While in the RFS phase azacitidine is associated with a [redacted] increase in utilities over CCR that is accompanied by a [redacted] increment in costs. As for the Relapse/PD phase these

figures are [redacted] and [redacted]. Consistent with our critique in *Section 5.3.6.1* these results may be biased as they are derived from extrapolations of RFS data based on parametric functions that invalidly imply proportional hazards, and which determine the length of Relapse/PD phase given that OS duration was determined separately from PFS and RFS duration. Further, as discussed in the previous section and in *Sections 5.3.8.1* and *5.3.8.2*, the estimation of costs in RFS is incorrect due to invalid account for the number of treatment cycles and counting two cycles instead of one in the first model cycle.

Table 55: Summary of QALYs by health state in company base case

Health state	QALY azacitidine	QALY CCR	Increment	Absolute increment	% absolute increment
RFS	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
PFS	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
PD	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Total	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Key: QALY, quality-adjusted life-year, CCR; conventional chemotherapy regimens, RFS; relapse free survival, PFS; progression free survival PD, progressive disease
Source: Celgene Submission, Table 56, p.146

Table 56: Summary of costs by health state in company base case

Health state	Cost azacitidine	Cost CCR	Increment	Absolute increment	% absolute increment
RFS	[redacted]	£6,503	[redacted]	[redacted]	[redacted]
PFS	[redacted]	£22,235	[redacted]	[redacted]	[redacted]
PD	[redacted]	£6,260	[redacted]	[redacted]	[redacted]
Terminal care	[redacted]	£5,609	[redacted]	[redacted]	[redacted]
Total	[redacted]	£40,608	[redacted]	[redacted]	[redacted]

Key: CCR; conventional chemotherapy regimens, RFS; relapse free survival, PFS; progression free survival PD, progressive disease
Source: Celgene Submission, Table 57, p.147

In addition, the PD costs assume healthcare resource utilisation associated with BSC, whereas, as discussed in *Section 5.3.6.2*, estimates of relative overall survival effectiveness and thus duration of the PD phase were estimated on AML-AZA-001 study data that were affected by the use of subsequent AML active therapy, to a larger extent in the azacitidine than the CCR arm.

5.3.10 Sensitivity analyses

When sampling uncertainty in the parameter values used in the base case were accounted for, the probabilistic sensitivity analysis resulted in an ICER of £17,423 per QALY, which is lower than the base case deterministic ICER of £20,648 per QALY. Taken at face value these different results between the deterministic and PSA results would suggests that there are important nonlinearities in the model that make PSA results more adequate estimates of cost-effectiveness than the deterministic values. However, the ERG found a number of problems as discussed before in this critique, which invalidate both the deterministic and PSA results. The ERG has corrected some errors in the implementation of the model and estimates of some parameter values populating it, as presented below in *Table 57*.

Despite the limitations identified in the model from which Celgene derived their base case cost-effectiveness results, its analysis of the most influential individual parameter values provides guidance as to which elements may need special consideration in assessing the validity of results. As described in *Section 5.2.9.1* (*page 92*), ICER results are most sensitive to variation in the costs of drug administration in CCR, followed by the hazard ratio of OS, with acquisition costs of AZA, administration costs of AZA and the HR of RFS being the fourth, fifth and sixth most influential parameters. In these parameters ERG found flaws either in model implementation or parameter estimation, as discussed above. To illustrate the importance of this, Celgene reports that a 20% increase in the value of these parameters is associated with an increase of their base case results of 155%, 44%, 113%, 98% and 78% (Celgene submission, Table 61, p. 153).

Of particular interest is the subgroup analysis of patients who would be candidates for LDAC under the control situation. In this group Celgene's scenario analysis is reported to result in an ICER of £25,136 per QALY with IPCW adjustment for treatment switching in the control arm. However, some 28% of the azacitidine arm subjects received subsequent active treatment but their health outcomes and costs were not adjusted for the effect of subsequent lines of treatment. This caveat is also relevant for the comparison of patients pre-selected for BSC and for IC controls, where according to results presented by Celgene, azacitidine generates more QALYs and has lower costs relative to BSC and IC. However, as explained next, the results of this subgroup analyses did not collectively pass a fundamental validation test.

Celgene provides the results of a scenario where the distribution of patients by pre-selected CCR treatment is that observed in clinical practice from HMRN registry data ([redacted] IC, [redacted] LDAC, [redacted] BSC) instead of the original trial case mix (18%, 64% and 18%, respectively). The ICERs change from £20,648 to -£57,756 per QALY, i.e., to the result that azacitidine is both more effective and less costly than CCR. Although an improvement in cost-effectiveness of azacitidine is to be expected when applying a distribution of the patient population with a larger proportion of patients in poorer health condition and thus eligible to receive BSC under CCR, as in AZA-AML-001 the largest detectable difference was found precisely in those patients, these are unexpectedly large results. ERG replicated these results to a small degree of discrepancy (-£57,968 versus -£57,656 per QALY), using the weighted average method described by Celgene in its submission (Celgene submission, Section 5.8.3, p. 155). However when ERG tried to replicate the base case results using the same method we obtained and ICER of -£17,960 per QALY (azacitidine dominant), quite a different result from the base case model result of £20,648 per QALY. This suggests that the method used by Celgene to calculate an alternative ICER based on HMRN data may not be compared with that used in the base case. It must be noted that this discrepancy is despite both the subgroup analysis by CCR pre-selected treatment and the analysis of the overall sample using the same OS effectiveness estimates, i.e., the exponential extrapolation using the adjusted IPCW HR estimate of [redacted].

The inability to replicate the base case results using the weighted average method described in the Celgene submission poses a severe limitation to ERG's ability to correct the most important flaws of the model submitted by Celgene. This is because the error incurred by the company in implementing the number of treatment cycles described in *Section 5.3.8.1* (*page 122*) may only be corrected, within their model, by separately calculating costs and QALYs for each CCR pre-selected treatment subgroup and combining results using weighted

averages of costs and of QALYs using the patient distributions in AZA-AML-001, in the same way as the company did for their sensitivity analysis of the HMRN patient population. Our corrected results presented in *Section 6* (*page 130*) are subject to this caveat.

5.3.11 Model validation and face validity check

5.3.11.1 Internal validity

We conducted two assessments of internal validity of the model submitted by Celgene. First, we compared the mean number of treatment cycles produced by the model in the azacitidine arm with the reported mean number of cycles in AZA-AML-001.[45] As discussed in *Section 5.3.8.1* we found that the model underestimated the mean number of treatment cycles, and this was due to an error of implementation in that the maximum number of treatment cycles allowed by the model was 8 (after rounding 8.8 to the lowest integer), thus overlooking the fact that a proportion of the modelled cohort would have made a transition to Relapse/PD and therefore be receiving no treatment within those first 8 cycles. Similar but more severe underestimation problems applied to the active treatments in the CCR arm.

The second validity check compared the overall survival outputs from the model with the respective outcome in the original trial report, for the azacitidine arm. In this case the model overestimated the observed data; the median overall survival was approximately 11.5 in the model base case (exponential distribution) versus 10.4 months in the Kaplan-Meier OS curve in AZA-AML-001 (Figure 1A of Dombret et al. 2015[45]). Incidentally, the model output was closer to the median OS obtained by censoring data at switch to subsequent AML therapy, i.e., 12.1 months (Figure 1B of Dombret et al. 2015[45]).

5.3.11.2 Model implementation checks

We conducted a list of model checks, including black-box tests (varying inputs and checking for the anticipated impact on outputs) and checking individual formulae in the model. We highlight below problems identified by this and a subsequent process that sought to replicate model results in the submission.

In calculating the costs of drug administration, monitoring tests and transfusions corresponding to stable disease ("Model CCR" and "Model AZA" columns AB:AD), the company subtracted a portion of these costs equal either to the differences between the proportion of patients in the modelled cohort who were in Non-remission and the proportion of patients who were in remission at each cycle, or to zero, whichever of the two quantities was higher. This was found to have no rational basis and was not explained nor described in the submission.

The model was also subject to verification tests, where individual parameters' values in the model were varied and results compared to *a priori* expectations. Further, the ERG identified an error in the formulae referencing the Kaplan-Meier curves for PFS and RFS in the "KM" worksheet (DD:DI columns). This is correction was important in the light of ERG's preference for using Kaplan-Meier time to event curves to model the evolution in these outcomes, as described in the next section.

5.4 Exploratory and sensitivity analyses undertaken by the ERG

5.4.1 Corrected base case

The ERG identified a number of implementation errors in the model submitted by Celgene and corrected them, as shown in *Table 57*.

Table 57: Corrections to the implementation of Celgene's model

Celgene's model	ERG's corrections
The costs of drug administration, tests and transfusions in LDAC patients were estimated for 8 model cycles, as for azacitidine, instead of 6.1 cycles reported in the submission (AB, AC and AD in 'Model CCR').	Corrected to incorporate changes in the mean number of treatment cycles for LDAC patients
Transfusion costs pre-response in CCR arm were modelled using the corresponding costs for azacitidine group ('Model CCR'!AD23)	Corrected
Transfusion costs in cycles 2+ in CCR arm were modelled incorrectly: the cost of transfusion in remission state was assumed to be equal to the transfusion costs in patients with stable disease ('Model CCR'!AD24:153).	Corrected
Cytarabine daily dose for LDAC assumed mg/m² (resulting in 71.82 mg/day) instead of 39.9 mg/day reported in CSR. However, it has no effect on costs in the base case and has a minimal effect on ICER in scenario analyses for no wastage and wastage with 30% tolerance ('HRU costs'!C115:117).	39.9 mg/day of cytarabine in LDAC patients
Kaplan-Meier OS, PFS and RFS curves for the overall sample were incorrectly referenced to the IC curves in the KM worksheet (DD:DI columns)	Corrected
Costs of tests and transfusions in PD state were not modelled by Celgene	Corrected
Celgene assume the drug administration costs for IC patients after cycle 3	Corrected
"BSC only" patients are assumed to incur drug administration costs	Corrected
Wastage with 30% tolerance was coded incorrectly	Corrected
The number of treatment cycles for which drug administration, monitoring tests and transfusions costs were accounted was two in the first treatment cycle (AB23, AC23 and AD23 in Model AZA and Model CCR), in contrast to drug acquisition costs for which only one cycle was accounted in the first cycle (AA23 in Model AZA and Model CCR sheets).	Corrected

Key: CCR, conventional care regimen; CSR, clinical study report; ERG, Evidence Review Group; ICER, incremental cost-effectiveness ratio; LDAC, low-dose cytarabine; OS, overall survival; PFS, progression-free survival; RFS, relapse-free survival

5.4.2 ERG preferred base case

The corrections to the implementation were followed by a series of changes made to the model parameter values to reflect what ERG considers the best values to reflect current UK practice and model logic. The results of these changes are presented as a sequence of cumulative changes in *Section 6*.

5.4.2.1 Calibrating the number of treatment cycles

The maximum number of treatment cycles of azacitidine treatment was set to 19, in order to match the 8.80 mean number of cycles in the respective arm of AZA-AML-001.[45] Similarly the maximum number of treatment cycles of LDAC was set to 10 to match the respective mean reported in the trial of 6.1 cycles of treatment. The IC was set to a maximum of two cycles as specified in the trial protocol and reported in the submission.

5.4.2.2 Equalising costs of relapse and progressive disease across treatments

The frequency and amount of use of medical staff, monitoring and outpatient procedures and hospitalisations used by patients managed under BSC in the progressive disease/relapse state as estimated from clinical opinion surveyed by Celgene was applied to IC and LDAC preselected CCR patients and azacitidine patients that were in the progressive disease or relapse health state. *Appendix 3* details the costs used in PD/Relapse by ERG for its preferred base case analysis.

5.4.2.3 Adjusting overall survival in both arms for subsequent active treatment

The ERG set the OS curves to the Censor At Switch analysis mode in Celgene's model, keeping the exponential functional form adopted in Celgene's base-case. This choice was associated with a hazard ratio of 0.72, which corresponded to the analysis that was unadjusted for baseline covariates. Given the high ICERs that were obtained after the preceding revisions in *Sections 5.4.2.1* and *5.4.2.2*, the effect of adopting the smaller hazard that resulted from censored at switch analysis that adjusted for baseline covariates was investigated in exploratory analyses.

5.4.2.4 Fitting separate parametric survival curves to relapse-free survival and progression-free survival in each arm

Since finding the optimal fitting functional form for RFS and PFS is highly uncertain as discussed in this report, the ERG adopted the observed non-parametric Kaplan-Meier curves from the trial for these outcomes. The rationale for this choice is further strengthened by the fact that the Kaplan-Meier curves are almost completely observed by the end of the observation period at 37 months and extrapolation is not necessary. Neither is it obvious that adjusting those survival curves for any observed confounders is practicable or indeed desirable.

5.4.2.5 Adjusting overall survival for baseline covariates

The effect of using the OS hazard ratio estimate of 0.64 from the exponential model adjusted for baseline covariates, which found support in statistical tests conducted by the ERG, was investigated.

5.4.3 Exploratory analyses

The ERG sought to perform some exploratory assessment of the subgroup analysis by preselected CCR treatment, while acknowledging that for PFS and RFS outcomes, the sample sizes make subgroup-specific time to event data highly unreliable. Thus in these analyses subgroup specific differences in OS outcomes were allowed using censor-at-switch data, while keeping common PFS and RFS curves across the three subgroups.

5.5 Conclusions of the cost-effectiveness section

The ERG identified several issues with the company's submitted economic evaluation.

The model assumed that no patients would receive active treatment following discontinuation of first-line treatment. In the AZA-AML-001 trial underpinning the analysis, 29% of participants received active second-line treatment. Advice from clinical experts suggests that active second-line treatment is considered for some patients in the NHS.

The model assumed proportional hazards for all time-to-event outcomes, even though this was not supported for overall survival and relapse-free survival by results from the AZA-AML-001 trial.

Overall survival in the AZA arm was not adjusted for subsequent active treatment, resulting in an inconsistency between the modelled health outcomes and costs, since only the costs of best supportive care were modelled following azacitidine.

Implementation issues were identified in the model. The most significant of these was an error in the calculation of the duration of first-line treatment which resulted in an underestimate of the drug acquisition and administration costs in both arms.

The ERG also identified that there were significant differences in the cost associated with the Relapse/progressive disease state between the AZA and CCR arm, even though all patients (in both arms) are expected to be receiving BSC at this point.

6 Impact on the ICER of additional clinical and economic analyses undertaken by the ERG

The ERG's preferred base case ICER is £169,606 per QALY (see *Table 58* and *Table 59*).

Corrections to implementation errors in the model increased the ICER from the base case £20,648 to £62,518 per QALY (analysis A).

Analyses B to G are the additional changes made to reach the ERG's preferred base case. Of these, two independently lead to significant increases in the ICER: calibrating the number of treatment cycles to match the mean number of cycles in AZA-AML-001[45] increases the ICER to £131,698 per QALY; setting the costs of relapse/progressive disease equal across the arms increases the ICER to £159,352 per QALY. Two independently lead to reductions in the ICER: adjusting overall survival for treatment switching in both arms reduces the ICER to £47,482 per QALY; adjusting overall survival for baseline covariates reduces the ICER to £39,145 per QALY. Using Kaplan-Meier relapse-free survival has little impact on the ICER (£63,569 per QALY). Using Kaplan-Meier progression-free survival increases the ICER to £75,471 per QALY.

Table 58: Corrected base case and elements of ERG preferred base case

Analysis	Outcome	Azacitidine	CCR	Difference
Celgene base case	Costs	[redacted]	£40,608	[redacted]
	QALYs	[redacted]	0.637	[redacted]
	ICER (cost per QALY gained)			£20,648
A = Corrected base case[a]	Costs	[redacted]	£45,954	[redacted]
	QALYs	[redacted]	0.637	[redacted]
	ICER (cost per QALY gained)			£62,518
A + B = A **and** Calibrating number of treatment cycles[b]	Costs	[redacted]	£50,064	[redacted]
	QALYs	[redacted]	0.637	[redacted]
	ICER (cost per QALY gained)			£131,698
A + C = A **and** Using the same costs of Relapse/PD across treatments[c]	Costs	[redacted]	£68,688	[redacted]
	QALYs	[redacted]	0.637	[redacted]
	ICER (cost per QALY gained)			£159,352
A + D = A **and** Overall survival adjusted for treatment switching in both arms[d]	Costs	[redacted]	£52,225	[redacted]
	QALYs	[redacted]	0.728	[redacted]
	ICER (cost per QALY gained)			£47,482
A + E = A **and** Kaplan-Meier RFS curves for each trial arm[e]	Costs	[redacted]	£46,221	[redacted]
	QALYs	[redacted]	0.636	[redacted]
	ICER (cost per QALY gained)			£63,569
A + F = A **and** Kaplan-Meier PFS curves for each trial arm[e]	Costs	[redacted]	£45,753	[redacted]
	QALYs	[redacted]	0.635	[redacted]
	ICER (cost per QALY gained)			£75,471
A + G = A **and** Relative OS effects from adjusted parametric curves[f]	Costs	[redacted]	£36,028	[redacted]
	QALYs	[redacted]	0.391	[redacted]
	ICER (cost per QALY gained)			£39,145

Key: CCR, conventional care regimen; ERG, Evidence Review Group; ICER, incremental cost-effectiveness ratio; PD, progressive disease; PFS, progression-free survival; QALY, quality-adjusted life year; RFS, relapse-free survival

Notes: a, See *Table 57*; b, See *Section 5.4.2.1*; c, See *Section 5.4.2.2*; d, See *Section 5.4.2.3*; e, See *Section 5.4.2.4*; f, See *Section 5.4.2.5*

Table 59: Derivation of the ERG's preferred base case

Analysis[a]	Outcome	Azacitidine	CCR	Difference
Celgene base case	Costs	[redacted]	£40,608	[redacted]
	QALYs	[redacted]	0.637	[redacted]
	ICER (cost per QALY gained)			£20,648
A = Corrected base case	Costs	[redacted]	£45,954	[redacted]
	QALYs	[redacted]	0.637	[redacted]
	ICER (cost per QALY gained)			£62,518
A + B	Costs	[redacted]	£50,064	[redacted]
	QALYs	[redacted]	0.637	[redacted]
	ICER (cost per QALY gained)			£131,698
A + B + C	Costs	[redacted]	£72,798	[redacted]
	QALYs	[redacted]	0.637	[redacted]
	ICER (cost per QALY gained)			£238,674
A + B + C + D	Costs	[redacted]	£91,847	[redacted]
	QALYs	[redacted]	0.728	[redacted]
	ICER (cost per QALY gained)			£171,511
A + B + C + D + E	Costs	[redacted]	£92,676	[redacted]
	QALYs	[redacted]	0.727	[redacted]
	ICER (cost per QALY gained)			£174,205
A + B + C + D + E + F	Costs	[redacted]	£98,046	[redacted]
	QALYs	[redacted]	0.724	[redacted]
	ICER (cost per QALY gained)			£246,488
A + B + C + D + E + F + G = ERG preferred base case	Costs	[redacted]	£41,161	[redacted]
	QALYs	[redacted]	0.390	[redacted]
	ICER (cost per QALY gained)			£169,606

Key: CCR, conventional care regimens; ERG, Evidence Review Group; ICER, incremental cost-effectiveness ratio; QALY, quality-adjusted life year
Note: a, See *Table 58*

6.1 Exploratory analyses

Exploratory subgroup analyses by preselected CCR treatment using the changes A–F described in *Table 58* by the ERG produce ICERs above £100,000 per QALY for all subgroups (*Table 60*). An adjustment for baseline covariates, which is not reliable due to the small sample sizes available within each group, would be expected to reduce these figures but they would remain around the £100,000 per QALY value.

Table 60: Scenarios explored for subgroup analysis explored by ERG

Scenario			Pre-selected CCR therapy subgroup	Incremental		ICER (cost per QALY)[b]
Analysis	PFS and RFS	OS		Costs	QALYs	
Celgene	PH Gompertz and PH Weibull	IPCW applied to CCR arm for switching to azacitidine	IC	[redacted]	[redacted]	-£52,184
			LDAC	[redacted]	[redacted]	£25,136
			BSC	[redacted]	[redacted]	-£169,672
Celgene	PH Gompertz and PH Weibull	ITT	IC	[redacted]	[redacted]	-£85,266
			LDAC	[redacted]	[redacted]	£41,671
			BSC	[redacted]	[redacted]	-£50,300
ERG[a]	Kaplan-Meier	Censored at switch for any active AML treatment	IC	[redacted]	[redacted]	£210,767
			LDAC	[redacted]	[redacted]	£276,260
			BSC	[redacted]	[redacted]	£98,715
ERG[a]	Kaplan-Meier	ITT	IC	[redacted]	[redacted]	£122,722
			LDAC	[redacted]	[redacted]	£408,492
			BSC	[redacted]	[redacted]	£80,952

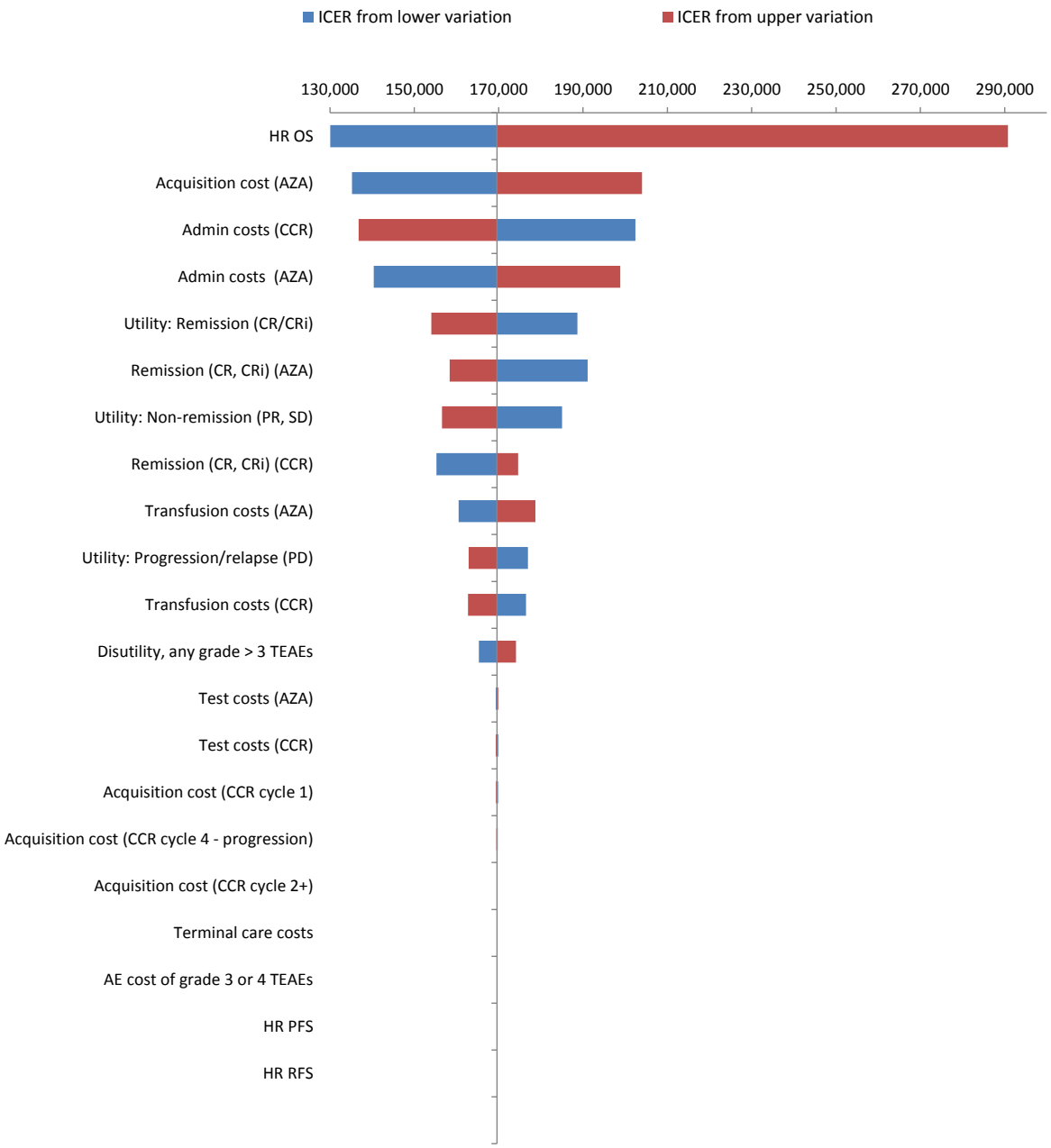
Key: AML, acute myeloid leukaemia; BSC, best supportive care; CCR, conventional care regimens; ERG, Evidence Review Group; IC, intensive chemotherapy; ICER, incremental cost-effectiveness ratio; IPCW, inverse probability of censoring weight; ITT, intention-to-treat; LDAC, low-dose cytarabine; OS, overall survival; PFS, progression-free survival; PH, proportional hazards; QALY, quality-adjusted life year; RFS, relapse-free survival

Notes: a, Includes corrections and changes as described in *Table 59* except for component 'G' (i.e., not including adjustment for baseline covariates); b, Negative ICERs indicate azacitidine is dominant

6.2 Univariate sensitivity analyses

The univariate sensitivity analysis with the base case preferred by ERG is presented in the tornado analysis of *Figure 23*; plausible variation of parameter values results in ICERs above £130,000 per QALY.

Figure 23: Tornado diagram of ERG's preferred base case deterministic analysis



Key: AZA, azacitidine; 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; RFS, relapse-free survival; SD, stable disease; TEAE, treatment emergent adverse event

6.3 Probabilistic sensitivity analysis

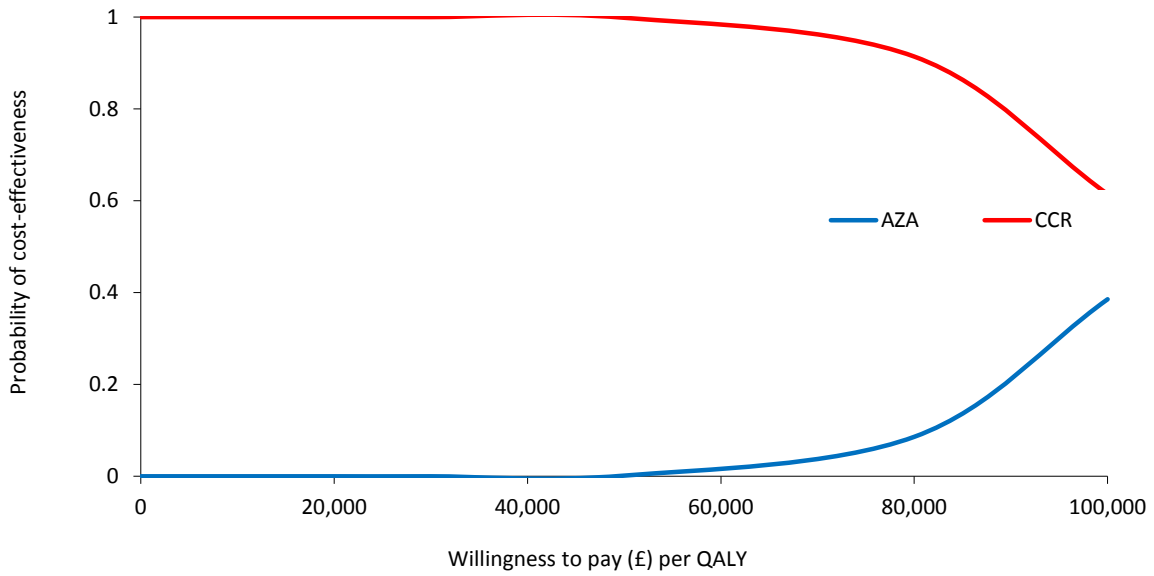
A probabilistic sensitivity analysis was conducted for the ERG's preferred base case (Table *61*). An ICER of £182,151 per QALY was obtained, which is similar to the deterministic ICER of £169,606 per QALY. *Figure 24* presents the cost-effectiveness acceptability curves from the probabilistic sensitivity analysis in the ERG's preferred base case. At a willingness to pay threshold of £100,000 the probability of azacitidine being cost-effective is less than 40%.

Table 61: Cost-effectiveness results for ERG's preferred base case probabilistic sensitivity analysis

Arm	Total			Incremental			ICER (cost per QALY)
	Costs	LYG	QALYs	Costs	LYG	QALYs	
CCR	£41,063	0.5478	0.3911	—	—	—	—
Azacitidine	[redacted]	0.8186	[redacted]	[redacted]	0.2708	[redacted]	£182,151

Key: CCR; conventional care regimens; ERG, Evidence Review Group; ICER, incremental cost-effectiveness ratio; LYG, life years gained; QALYs, quality-adjusted life years

Figure 24: Cost-effectiveness acceptability curves from ERG's preferred base case probabilistic sensitivity analysis



Key: AZA, azacitidine; CCR, conventional care regimens; ERG, Evidence Review Group; QALY, quality-adjusted life year

7 End of life

The NICE Guide to the methods of technology appraisal[41] indicates that while in the reference case all QALYs are regarded as being of equal weight, the Appraisal Committee can consider QALY weighting in the case of life-extending treatment at the end of life.

Celgene included an assessment of three criteria (all of which should be met for end of life consideration), and this is reproduced accompanied by ERG comments in *Table 62*.

Table 62: Assessment of end-of-life criteria

Criterion	Data available (Celgene)	ERG comment
The treatment is indicated for patients with a short life expectancy, normally less than 24 months	Median OS reported in the literature ranges between 1.5 months (aged >65 years) and 2 months (aged >55 years)[18, 19]	Median OS in AZA-AML-001 trial is 6.5 months without azacitidine treatment. Restricted mean survival at 30 months is estimated to be 10.55 months (*Appendix 4*). Azacitidine is also indicated for intermediate-2 and high-risk myelodysplastic syndromes, CMML with 10–29% marrow blasts without myeloproliferative disorder and AML with 20–30% blasts and multi-lineage dysplasia. The results of the AZA-001 trial in this population suggest median OS of 15.0 months without azacitidine treatment.[69]
There is sufficient evidence to indicate that the treatment offers an extension to life, normally of at least an additional 3 months, compared with current NHS treatment	Median OS based on the primary endpoint was 10.4 months in the azacitidine group and 6.5 months in the CCR group, providing an OS benefit of 3.8 months with azacitidine. As reported in Section 4.7 [of Celgene's submission], various pre-defined analyses demonstrated that treatment with azacitidine provided a statistically significant survival benefit versus CCR	Extension to life should be assessed considering differences in mean overall survival in addition to median OS. ERG analyses based on restricted mean survival at 30 months suggest an extension to life of 1.8–2.5 months (depending on how treatment switching is handled – see *Appendix 4*). The estimated improvement in restricted mean OS is greatest for patients pre-selected to BSC, although comparisons to individual CCR are subject to uncertainty.
The treatment is licensed or otherwise indicated for small patient populations	The estimated total population for all licensed indications in England is 3,354, consisting of 1,026 covered by the proposed new indication and 2,328 for all existing indications. See additional detail provided [in Celgene submission, Section 4.13.2, pp. 97–98]	No comment.

Key: AML, acute myeloid leukaemia; BSC, best supportive care; CCR, conventional care regimen; CMML, chronic myelomonocytic leukaemia; ERG, Evidence Review Group; OS, overall survival; NHS, National Health Service.

Source: Celgene submission, Table 28, p.97

The NICE guide also indicates that the Committee will need to be satisfied that the estimates of the extension to life are robust, and that the assumptions used in the reference case economic modelling are plausible, objective and robust.

The ERG considers that the estimates of extension to life are not robust. The company's estimate based on median overall survival difference is not robust as it is does not reflect the convergence of survival curves seen in AZA-AML-001. Overall survival benefit was also the primary endpoint of AZA-AML-001 and this did not reach statistical significance. Estimates of overall survival are affected by adjustments for baseline imbalances and for subsequent active treatment.

The ERG conducted additional analyses of the restricted mean overall survival at 30 months (see *Appendix 4*) and found that the survival gain was less than three months on average for azacitidine versus CCR.

The ERG also considers that the assumptions used in the company's economic modelling were not plausible or robust. In *Section 6* (*page 130*) the ERG show the results of correcting these assumptions and implementation errors. The ERG's preferred base case ICER is considered to be more plausible, objective and robust.

8 Overall conclusions

The ERG believes that by limiting the comparators in the decision problem to a combined CCR comparator, there is the possibility of azacitidine being recommended or not recommended inappropriately for certain patients (according to the most appropriate CCR for them). Although it is claimed that these patients cannot be reliably and objectively identified, they are at least sufficiently identifiable for CCR regimens to be assigned both in routine clinical practice and in the pivotal RCT. In the absence of high-quality clinical effectiveness data on the clinical value of azacitidine versus individual conventional care regimens, clinical decision making may be made more challenging if azacitidine is recommended for all elderly AML patients with >30% blasts.

Of the key changes made to the economic model by the ERG in arriving at their preferred base case (ICER £169,606 per QALY), the following might be considered to be differences of opinion between the company and the ERG:

- Adjustment of overall survival in the azacitidine and CCR arms for subsequent active treatment – the company only adjusted survival in the CCR arm (for subsequent treatment with azacitidine), while the ERG has adjusted survival in both arms for any active treatment for AML;
- Equalising costs per month spent in the Relapse/progressive disease state – the company assumed significant incremental costs in the CCR arm despite all patients in the model receiving best supportive care only in this state; their assumption is based on a questionnaire-based survey of clinicians, but expert opinion solicited by the ERG does not support this assumption;
- Replacing parametric survival models for relapse-free and progression-free survival with Kaplan-Meier curves to avoid imposing a proportional hazards assumption – the company used a proportional hazards assumption "for consistency" although they acknowledged it was in contradiction of the data.

The ERG considers that the changes made to the modelled treatment duration cannot be considered matters of opinion and are corrections to implementation errors. After correcting solely for these implementation issues, an ICER of £131,698 per QALY was obtained.

The base case analysis preferred by the ERG has the limitation that the sum of the cohort proportions in PFS and RFS exceeds the proportion alive by the third cycle in the model. Celgene included an adjustment in the model to eliminate such anomalies, but nevertheless this remains a deficiency of the model structure. The ERG has not attempted to correct this, due to the difficulties of fitting plausible parametric models to the PFS and RFS data (as discussed in this critique) and adjusting for baseline covariates, given the small samples available for analysis.

8.1 Implications for research

8.1.1 Clinical effectiveness

Further research is needed to establish the effectiveness of azacitidine in elderly AML patients with >30% blasts. This research should be powered to detect clinically meaningful improvements in survival between azacitidine and individual conventional care regimens.

The statistical analyses planned for the research should account for the likelihood of non-proportional hazards and for treatment switching, with adequate data collection to support multiple plausible statistical models.

8.1.2 Health-related quality of life (HRQoL)

This research should also collect HRQoL data measured using a generic (as opposed to condition-specific) and validated instrument, which allow outcomes to be valued using preferences from the general public (preferably EQ-5D) and is preferred for economic analyses. Significant efforts should be made to collect HRQoL data across all patients and across all time points to reflect the full range of quality of life experienced by patients.

8.1.3 Healthcare resource use

Further research (likely separate from the research above) is required to accurately estimate healthcare resource use in elderly AML patients with >30% blasts within the NHS. For conventional care regimens this should be based on routine clinical practice. For azacitidine, this may involve a pilot study, or, if azacitidine receives a positive recommendation from NICE, prospective collection of healthcare resource use data, and collection of data relating to the most clinically appropriate alternative treatment for each patient.

References

1. Lowenberg B, Downing JR, Burnett A. Acute myeloid leukemia. N Engl J Med. 1999;341(14):1051-62.
2. Sekeres MA, Stone RM, Zahrieh D, Neuberg D, Morrison V, De Angelo DJ, et al. Decision-making and quality of life in older adults with acute myeloid leukemia or advanced myelodysplastic syndrome. Leukemia. 2004;18(4):809-16.
3. Deschler B, Lubbert M. Acute myeloid leukemia: epidemiology and etiology. Cancer. 2006;107(9):2099-107.
4. Cancer Research UK. Leukaemia (AML) incidence by age. Available at: http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/leukaemia-aml/incidence#heading-Zero [Last accessed: October 2015].
5. Klepin HD, Rao AV, Pardee TS. Acute myeloid leukemia and myelodysplastic syndromes in older adults. J Clin Oncol. 2014;32(24):2541-52.
6. Burnett A, Milligan D, Prentice A, Goldstone A, McMullin M, Hills Rea. A comparison of low-dose cytarabine and hydroxyurea with or without all-trans retinoic acid for acute myeloid leukemia and high-risk myelodysplastic syndrome in patients not considered fit for intensive treatment. Cancer. 2007;109:114-24.
7. Juliusson G, Lazarevic V, Horstedt AS, Hagberg O, Hoglund M, Swedish Acute Leukemia Registry G. Acute myeloid leukemia in the real world: why population-based registries are needed. Blood. 2012;119(17):3890-9.
8. Lang K, Earle CC, Foster T, Dixon D, Van Gool R, Menzin J. Trends in the treatment of acute myeloid leukaemia in the elderly. Drugs Aging. 2005;22(11):943-55.
9. Oran B, Weisdorf DJ. Survival for older patients with acute myeloid leukemia: a population-based study. Haematologica. 2012;97(12):1916-24.
10. Meyers J, Yu Y, Kaye JA, Davis KL. Medicare fee-for-service enrollees with primary acute myeloid leukemia: an analysis of treatment patterns, survival, and healthcare resource utilization and costs. Appl Health Econ Health Policy. 2013;11(3):275-86.
11. Haematological Malignancy Research Network. Incidence. Available at: https://www.hmrn.org/statistics/incidence [Last accessed: November 2015].
12. Office for National Statistics. Population Estimates for UK, England and Wales, Scotland and Northern Ireland, Mid-2014. Available at: http://ons.gov.uk/ons/publications/re-reference-tables.html?edition=tcm%3A77-368259 [Last accessed: November 2015]. 2015.
13. Office for National Statistics. Mid-year population estimates for the UK 2014. Available at: http://ons.gov.uk/ons/rel/pop-estimate/population-estimates-for-uk--england-and-wales--scotland-and-northern-ireland/mid-2014/mid-year-population-estimates-for-the-uk-2014.html [Last accessed: November 2015]. 2015.
14. Vardiman JW, Thiele J, Arber DA, Brunning RD, Borowitz MJ, Porwit A, et al. The 2008 revision of the World Health Organization (WHO) classification of myeloid neoplasms and acute leukemia: rationale and important changes. Blood. 2009;114(5):937-51.
15. Vey N. Targeting age-related changes in the biology of acute myeloid leukemia: is the patient seeing the progress? Interdiscip Top Gerontol. 2013;38:73-84.

16. Sherman AE, Motyckova G, Fega KR, Deangelo DJ, Abel GA, Steensma D, et al. Geriatric assessment in older patients with acute myeloid leukemia: a retrospective study of associated treatment and outcomes. Leuk Res. 2013;37(9):998-1003.

17. Klepin HD, Geiger AM, Tooze JA, Kritchevsky SB, Williamson JD, Pardee TS, et al. Geriatric assessment predicts survival for older adults receiving induction chemotherapy for acute myelogenous leukemia. Blood. 2013;121(21):4287-94.

18. Taylor PR, Reid MM, Stark AN, Bown N, Hamilton PJ, Proctor SJ. De novo acute myeloid leukaemia in patients over 55-years-old: a population-based study of incidence, treatment and outcome. Northern Region Haematology Group. Leukemia. 1995;9(2):231-7.

19. Medeiros B, Satram-Hoang S, Hurst D, Hoang K, Momin F, Reyes C. Big data analysis of treatment patterns and outcomes among elderly acute myeloid leukemia patients in the United States. Ann Hematol. 2015;94:1127-38.

20. Celgene Ltd. Date on File: Haematological Malignancy Research Network: Clinical management and outcome in acute myeloid leukaemia patients. 2015.

21. National Cancer Institute. Surveillance, Epidemiology, and End Results Program. Actute myeloid leukemia, 5-year relative survival (percent) 2003-2009 by age at diagnosis. Available at: http://seer.cancer.gov/archive/csr/1975_2010/browse_csr.php?sectionSEL=13&pageSEL=sect_13_table.16.html#table3 [Last accessed October 2015].

22. Juliusson G, Antunovic P, Derolf A, Lehmann S, Mollgard L, Stockelberg D, et al. Age and acute myeloid leukemia: real world data on decision to treat and outcomes from the Swedish Acute Leukemia Registry. Blood. 2009;113(18):4179-87.

23. 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;30(36):4515-23.

24. Oliva EN, Nobile F, Alimena G, Ronco F, Specchia G, Impera S, et al. Quality of life in elderly patients with acute myeloid leukemia: patients may be more accurate than physicians. Haematologica. 2011;96(5):696-702.

25. Stalfelt AM, Brodin H, Pettersson S, Eklof A. The final phase in acute myeloid leukaemia (AML). A study on bleeding, infection and pain. Leuk Res. 2003;27(6):481-8.

26. Bevans M, Sternberg EM. Caregiving burden, stress, and health effects among family caregivers of adult cancer patients. JAMA. 2012;307(4):398-403.

27. British Committee for Standards in Haematology, Milligan DW, Grimwade D, Cullis JO, Bond L, Swirsky D, et al. Guidelines on the management of acute myeloid leukaemia in adults. Br J Haematol. 2006;135(4):450-74.

28. Dohner H, Estey EH, Amadori S, Appelbaum FR, Buchner 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;115(3):453-74.

29. European Medicines Agency. Guideline on the evaluation of anticancer medicinal products in man CPMP/EWP/205/95/Rev.3/Corr.2. 2006.

30. National Comprehensive Cancer Network. NCCN guidelines version 1.2015 acute myeloid leukemia. 2015.

31. Ossenkoppele G, Lowenberg B. How I treat the older patient with acute myeloid leukemia. Blood. 2015;125(5):767-74.

32. Burnett AK, Russell NH, Hunter AE, Milligan D, Knapper S, Wheatley K, et al. Clofarabine doubles the response rate in older patients with acute myeloid leukemia but does not improve survival. Blood. 2013;122(8):1384-94.

33. Estey E, de Lima M, Tibes R, Pierce S, Kantarjian H, Champlin R, et al. Prospective feasibility analysis of reduced-intensity conditioning (RIC) regimens for hematopoietic stem cell transplantation (HSCT) in elderly patients with acute myeloid leukemia (AML) and high-risk myelodysplastic syndrome (MDS). Blood. 2007;109(4):1395-400.

34. National Institute for Health and Care Excellence. TA270. Decitabine for the treatment of acute myeloid leukaemia (terminated appraisal). Available at: http://www.nice.org.uk/guidance/ta270 [Last accessed: November 2015]. 2012.

35. Howell DA, Wang HI, Roman E, Smith AG, Patmore R, Johnson MJ, et al. Variations in specialist palliative care referrals: findings from a population-based patient cohort of acute myeloid leukaemia, diffuse large B-cell lymphoma and myeloma. BMJ supportive & palliative care. 2015;5(5):496-502.

36. National Institute for Health and Care Excellence. Final Scope: Azacitidine for treating acute myeloid leukaemia with more than 30% bone marrow blasts. 2014.

37. Celgene Corporation. Clincal Study Report: AZA-AML-001 A phase 3, multicenter, randomized, open-label, study of azacitidine (Vidaza) versus conventional care regimens for the treatment of older subjects with newly diagnosed acute myeloid leukemia. 2014.

38. National Audit Office. Department of Health, NHS England and Public Health England. Progress in improving cancer services and outcomes in England. Available at https://www.nao.org.uk/wp-content/uploads/2015/01/Progress-improving-cancer-services-and-outcomes-in-England.pdf [Last accessed October 2015]. 2015.

39. Orphanet. Orphanet report series - list of rate diseases and synonyms listed in alphabetical order. Available at http://www.orpha.net/orphacom/cahiers/docs/GB/List_of_rare_diseases_in_alphabetical_order.pdf [Last accessed: October 2015]. 2015.

40. Department of Health. National Cancer Patient Experience Survey Programme. 2010 National Survey Report. Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/216682/dh_122520.pdf [Last accessed: October 2015]. 2010.

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

42. Karon M, Sieger L, Leimbrock S, Finklestein JZ, Nesbit ME, Swaney JJ. 5-Azacytidine: a new active agent for the treatment of acute leukemia. Blood. 1973;42(3):359-65.

43. Vogler WR, Miller DS, Keller JW. 5-Azacytidine (NSC 102816): a new drug for the treatment of myeloblastic leukemia. Blood. 1976;48(3):331-7.

44. McCormack SE, Warlick ED. Epigenetic approaches in the treatment of myelodysplastic syndromes: clinical utility of azacitidine. OncoTargets and therapy. 2010;3:157-65.

45. 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;126(3):291-9.

46. Lao Z, Yiu R, Wong GC, Ho A. Treatment of elderly patients with acute myeloid leukemia with azacitidine results in fewer hospitalization days and infective

complications but similar survival compared with intensive chemotherapy. Asia-Pacific journal of clinical oncology. 2015;11(1):54-61.

47. Vardiman JW, Harris NL, Brunning RD. The World Health Organization (WHO) classification of the myeloid neoplasms. Blood. 2002;100(7):2292-302.

48. Kantarjian HM, Thomas XG, Dmoszynska A, Wierzbowska A, Mazur G, Mayer J, et al. Multicenter, randomized, open-label, phase III trial of decitabine versus patient choice, with physician advice, of either supportive care or low-dose cytarabine for the treatment of older patients with newly diagnosed acute myeloid leukemia. J Clin Oncol. 2012;30(21):2670-7.

49. Osoba D, Rodrigues G, Myles J, Zee B, Pater J. Interpreting the significance of changes in health-related quality-of-life scores. J Clin Oncol. 1998;16(1):139-44.

50. Feng C, Wang H, Tu XM. Power Loss of Stratified Log-Rank Test in Homogeneous Samples. International Journal of Quality, Statistics, and Reliability. 2010;2010:4.

51. Latimer N, Abrams K. NICE DSU Technical Support Document 16: Adjusting survival time estimates in the presence of treatment switching. Available at: http://www.nicedsu.org.uk/TSD16_Treatment_Switching.pdf [Last accessed: November 2015]. 2014.

52. Cole SR, Hernan MA. Adjusted survival curves with inverse probability weights. Comput Methods Programs Biomed. 2004;75(1):45-9.

53. Howe CJ, Cole SR, Chmiel JS, Munoz A. Limitations of Inverse Probability-of-Censoring Weights in Estimating Survival in the Presence of Strong Selection Bias American Journal of Epidemiology. 2011;173(5):569-77.

54. Dufoir T, Saux MC, Terraza B, Marit G, Guessard S, Foulon G, et al. Comparative cost of allogeneic or autologous bone marrow transplantation and chemotherapy in patients with acute myeloid leukaemia in first remission. Bone Marrow Transplantation. 1992;10(4):323-9.

55. Huang BT, Wang Y, Du QF, Yang J, Yu J, Zeng QC, et al. Analysis of efficacy and cost-effectiveness of high-dose arabinoside versus daunorubicin chemotherapy in older adult patients with acute myeloid leukemia by cytogenetic risk profile: Retrospective review from China. International Journal of Hematology. 2011;93(4):474-81.

56. Kulikov A, Yagudina R, Misikova B. Pharmacoeconomic evaluation of acute myeloid leukemia and mds syndromes (intermediate and high risk) treatment with azacitidine in the russian federation. Value in Health. 2012;15 (7):A423.

57. 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.

58. Davies A, Briggs A, Schneider J, Levy A, Ebeid O, al e. The Ends Justify the Mean: Outcome Measures for Estimating the Value of New Cancer Therapies. 2012.

59. Proskorovsky I, Lewis P, Williams CD, Jordan K, Kyriakou C, Ishak J, et al. Mapping EORTC QLQ-C30 and QLQ-MY20 to EQ-5D in patients with multiple myeloma. Health Qual Life Outcomes. 2014;12:35.

60. McKenzie L, van der Pol M. Mapping the EORTC QLQ C-30 onto the EQ-5D instrument: the potential to estimate QALYs without generic preference data. Value Health. 2009;12(1):167-71.

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

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

63. NHS Blood and Transplant. Price List 2015/16; Code BC001; http://hospital.blood.co.uk/media/27457/price_list_2015-16.pdf, accessed on: 15/11/15.

64. NHS Blood and Transplant. Price List 2015/16; Code BC04; http://hospital.blood.co.uk/media/27457/price_list_2015-16.pdf, accessed on: 15/11/15.

65. Drummond MF, Jefferson TO. Guidelines for authors and peer reviewers of economic submissions to the BMJ. The BMJ Economic Evaluation Working Party. Bmj. 1996;313(7052):275-83.

66. Philips Z, Ginnelly L, Sculpher M, Claxton K, Golder S, Riemsma R, et al. Review of guidelines for good practice in decision-analytic modelling in health technology assessment. Health Technol Assess. 2004;8(36):iii-iv, ix-xi, 1-158.

67. Burnham KP, Anderson DR. Multimodel inference - understanding AIC and BIC in model selection. Sociol Method Res. 2004;33(2):261-304.

68. Kass RE, Wasserman L. A Reference Bayesian Test for Nested Hypotheses and Its Relationship to the Schwarz Criterion. J Am Stat Assoc. 1995;90(431):928-34.

69. Fenaux P, Mufti GJ, Hellstrom-Lindberg E, Santini V, Finelli C, Giagounidis A, et al. Efficacy of azacitidine compared with that of conventional care regimens in the treatment of higher-risk myelodysplastic syndromes: a randomised, open-label, phase III study. The Lancet Oncology. 2009;10(3):223-32.

List of appendices

Appendix 1: Additional clinical effectiveness results from AZA-AML-001

Appendix 2: Statistical analysis of individual patient data for overall survival in AZA-AML-001

Appendix 3: Resource use values applied by ERG to Relapse/PD phase across all arms for calculating their preferred base case analysis

Appendix 4: Estimation of restricted mean overall survival in AZA-AML-001

Azacitidine for treating acute myeloid leukaemia with more than 30% bone marrow blasts

Addendum

This addendum updates and replaces Section 6 “Impact on the ICER of additional clinical and economic analyses undertaken by the ERG”.

An error was identified in the ERG’s estimate of the baseline overall survival curve adjusted for baseline covariates. This has been corrected and the effect on the results has been reflected in the new Section 6. The ERG has also made certain clarifications within the section and has conducted additional exploratory analyses.

6 Impact on the ICER of additional clinical and economic analyses undertaken by the ERG

The ERG's preferred base case ICER is £273,308 per QALY (see *Table 58* and *Table 59*).

Corrections to implementation errors in the model increased the ICER from the base case £20,648 to £62,518 per QALY (analysis A).

Analyses B to G are the additional changes made to reach the ERG's preferred base case. Of these, two independently lead to significant increases in the ICER: calibrating the number of treatment cycles to match the mean number of cycles in AZA-AML-001[45] increases the ICER to £131,698 per QALY (analysis B); setting the costs of relapse/progressive disease equal across the arms increases the ICER to £159,352 per QALY (analysis C).

The primary focus of analysis D is to change the way overall survival is modelled by censoring for treatment switching in both arms. As a side effect (due to the model wiring) this also results in changes to the modelling of relapse-free and progression-free survival, again to censor for treatment switching in both arms. The effect of this analysis is to reduce the ICER to £47,482 per QALY.

Analysis E replaces the parametric proportional hazards progression-free survival curves with Kaplan-Meier curves and increases the ICER to £75,471 per QALY.

Using Kaplan-Meier curves for relapse-free survival (analysis F) has little impact on the ICER (£63,569 per QALY).

Adjusting overall survival for treatment switching (censoring at switch in both arms) and baseline covariates (analysis G) increases the ICER to £65,188 per QALY. The reason the ICER for analysis G is higher than the ICER for analysis D is that analysis G does not have the side effects on relapse-free and progression-free survival and so azacitidine patients spend longer in the progressive disease model state with high costs and low utility.

Table 58: Corrected base case and elements of ERG preferred base case

Analysis	Outcome	Azacitidine	CCR	Difference
Celgene base case	Costs	[redacted]	£40,608	[redacted]
	QALYs	[redacted]	0.637	[redacted]
	ICER (cost per QALY gained)			£20,648
A = Corrected base case[a]	Costs	[redacted]	£45,954	[redacted]
	QALYs	[redacted]	0.637	[redacted]
	ICER (cost per QALY gained)			£62,518
A + B = A **and** Calibrating number of treatment cycles[b]	Costs	[redacted]	£50,064	[redacted]
	QALYs	[redacted]	0.637	[redacted]
	ICER (cost per QALY gained)			£131,698
A + C = A **and** Using the same costs of Relapse/PD across treatments[c]	Costs	[redacted]	£68,688	[redacted]
	QALYs	[redacted]	0.637	[redacted]
	ICER (cost per QALY gained)			£159,352
A + D = A **and** Overall survival adjusted for treatment switching in both arms[d]	Costs	[redacted]	£52,225	[redacted]
	QALYs	[redacted]	0.728	[redacted]
	ICER (cost per QALY gained)			£47,482
A + E = A **and** Kaplan-Meier RFS curves for each trial arm[e]	Costs	[redacted]	£46,221	[redacted]
	QALYs	[redacted]	0.636	[redacted]
	ICER (cost per QALY gained)			£63,569
A + F = A **and** Kaplan-Meier PFS curves for each trial arm[e]	Costs	[redacted]	£45,753	[redacted]
	QALYs	[redacted]	0.635	[redacted]
	ICER (cost per QALY gained)			£75,471
A + G = A **and** Relative OS effects from adjusted parametric curves[f]	Costs	[redacted]	£44,818	[redacted]
	QALYs	[redacted]	0.622	[redacted]
	ICER (cost per QALY gained)			£65,188

Key: CCR, conventional care regimen; ERG, Evidence Review Group; ICER, incremental cost-effectiveness ratio; PD, progressive disease; PFS, progression-free survival; QALY, quality-adjusted life year; RFS, relapse-free survival

Notes: a, See *Table 57*; b, See *Section 5.4.2.1*; c, See *Section 5.4.2.2*; d, See *Section 5.4.2.3*, also note that PFS and RFS are adjusted as a side effect; e, See *Section 5.4.2.4*; f, See *Section 5.4.2.5*

Table 59: Derivation of the ERG's preferred base case

Analysis[a]	Outcome	Azacitidine	CCR	Difference
Celgene base case	Costs	[redacted]	£40,608	[redacted]
	QALYs	[redacted]	0.637	[redacted]
	ICER (cost per QALY gained)			£20,648
A = Corrected base case	Costs	[redacted]	£45,954	[redacted]
	QALYs	[redacted]	0.637	[redacted]
	ICER (cost per QALY gained)			£62,518
A + B	Costs	[redacted]	£50,064	[redacted]
	QALYs	[redacted]	0.637	[redacted]
	ICER (cost per QALY gained)			£131,698
A + B + C	Costs	[redacted]	£72,798	[redacted]
	QALYs	[redacted]	0.637	[redacted]
	ICER (cost per QALY gained)			£238,674
A + B + C + D	Costs	[redacted]	£91,847	[redacted]
	QALYs	[redacted]	0.728	[redacted]
	ICER (cost per QALY gained)			£171,511
A + B + C + D + E	Costs	[redacted]	£92,676	[redacted]
	QALYs	[redacted]	0.727	[redacted]
	ICER (cost per QALY gained)			£174,205
A + B + C + D + E + F	Costs	[redacted]	£98,046	[redacted]
	QALYs	[redacted]	0.724	[redacted]
	ICER (cost per QALY gained)			£246,488
A + B + C + D + E + F + G = ERG preferred base case	Costs	[redacted]	£71,138	[redacted]
	QALYs	[redacted]	0.621	[redacted]
	ICER (cost per QALY gained)			£273,308

Key: CCR, conventional care regimens; ERG, Evidence Review Group; ICER, incremental cost-effectiveness ratio; QALY, quality-adjusted life year
Note: a, See *Table 58*

6.1 Exploratory analyses

The ERG explored the extreme case that no healthcare costs are incurred in the Relapse/PD state, which resulted in an ICER of £73,953 per QALY.

In another scenario, in which no inpatient hospitalisations occur in the Relapse/PD state, the resulting ICER was £105,611 per QALY. This scenario may be relevant if 100% of inpatient days in the Relapse/PD phase are incurred due to terminal care; it seems plausible to ERG

that the cost estimate of terminal care of £5,705 in the Celgene model could already account for most hospital costs in PD, and therefore assuming zero inpatient costs may be a plausible scenario.

The base case analysis preferred by ERG included costs of monitoring tests and transfusions for the whole duration of the Remission and Non-remission phases of the model (as well as for the Relapse/PD phase). This was viewed as a correction of Celgene's analysis, which only measured these costs up to the time the patient stopped treatment. The ERG explored the effects of adopting Celgene's assumption: the ICER changed from £273,308 to £260,190 per QALY. Celgene's assumption is unlikely to be correct with respect to transfusions as acknowledged by Celgene's own definition of BSC (Source: Celgene submission, Section 4.3.5, p. 49, and Section 5.2.3, p. 108):

> *BSC: Including but is not limited to red cell or whole blood transfusions, fresh frozen plasma transfusions, platelet transfusions, antibiotic and/or antifungal therapy, and nutritional support). This is continued until death.*

Other scenarios included revising the costs of monitoring tests and transfusions during Relapse/PD to values estimated by a clinical expert consulted by ERG, whereby two units each of red blood cell transfusions and adult doses of platelet transfusions are given on average per 4 week cycle, whilst no bone marrow aspirates or biopsies nor extractions for cytogenetic testing, four blood tests, and two each of peripheral blood smears and serum blood chemistry are given during PD. The resulting ICER was £257,211 per QALY.

Exploratory subgroup analyses by preselected CCR treatment using the changes A–F described in *Table 58* by the ERG produce ICERs above £100,000 per QALY for all subgroups (*Table 60*). Exploratory subgroup analyses were also conducted by preselected CCR treatment using changes A, B and D–F (i.e., leaving in place Celgene's assumptions regarding costs in Relapse/PD), with the result that for patients preselected to intensive chemotherapy an ICER of £73,728 per QALY was obtained, while for other patients the ICER remained over £100,000 per QALY. An adjustment for baseline covariates, which is not reliable due to the small sample sizes available within each group, would be expected to increase the ICERs overall.

Table 60: Scenarios explored for subgroup analysis explored by ERG

Scenario			Pre-selected CCR therapy subgroup	Incremental		ICER (cost per QALY)[a]
Analysis	**PFS and RFS**	**OS**		**Costs**	**QALYs**	
Celgene, adjusted for subsequent therapies	Exponential and Weibull	IPCW applied to CCR arm for switching to azacitidine	IC	[redacted]	[redacted]	-£52,184
	Gompertz and Weibull		LDAC	[redacted]	[redacted]	£25,136
	Exponential and Weibull		BSC	[redacted]	[redacted]	-£169,672
Celgene, unadjusted for subsequent therapies	PH Gompertz and PH Weibull	Exponential	IC	[redacted]	[redacted]	-£85,266
		Gompertz	LDAC	[redacted]	[redacted]	£41,671
		Exponential	BSC	[redacted]	[redacted]	-£50,300
ERG[b]	Kaplan-Meier	Exponential, censored at switch for any active AML treatment	IC	[redacted]	[redacted]	£352,918
			LDAC	[redacted]	[redacted]	£282,589
			BSC	[redacted]	[redacted]	£152,093
ERG[b,c]	Kaplan-Meier	Exponential, censored at switch for any active AML treatment	IC	[redacted]	[redacted]	£73,728
			LDAC	[redacted]	[redacted]	£131,349
			BSC	[redacted]	[redacted]	£135,230
ERG[b]	Kaplan-Meier	ITT, Kaplan-Meier	IC	[redacted]	[redacted]	£414,304
			LDAC	[redacted]	[redacted]	£500,493
			BSC	[redacted]	[redacted]	£137,449

Key: AML, acute myeloid leukaemia; BSC, best supportive care; CCR, conventional care regimens; ERG, Evidence Review Group; IC, intensive chemotherapy; ICER, incremental cost-effectiveness ratio; IPCW, inverse probability of censoring weight; ITT, intention-to-treat; LDAC, low-dose cytarabine; OS, overall survival; PFS, progression-free survival; PH, proportional hazards; QALY, quality-adjusted life year; RFS, relapse-free survival

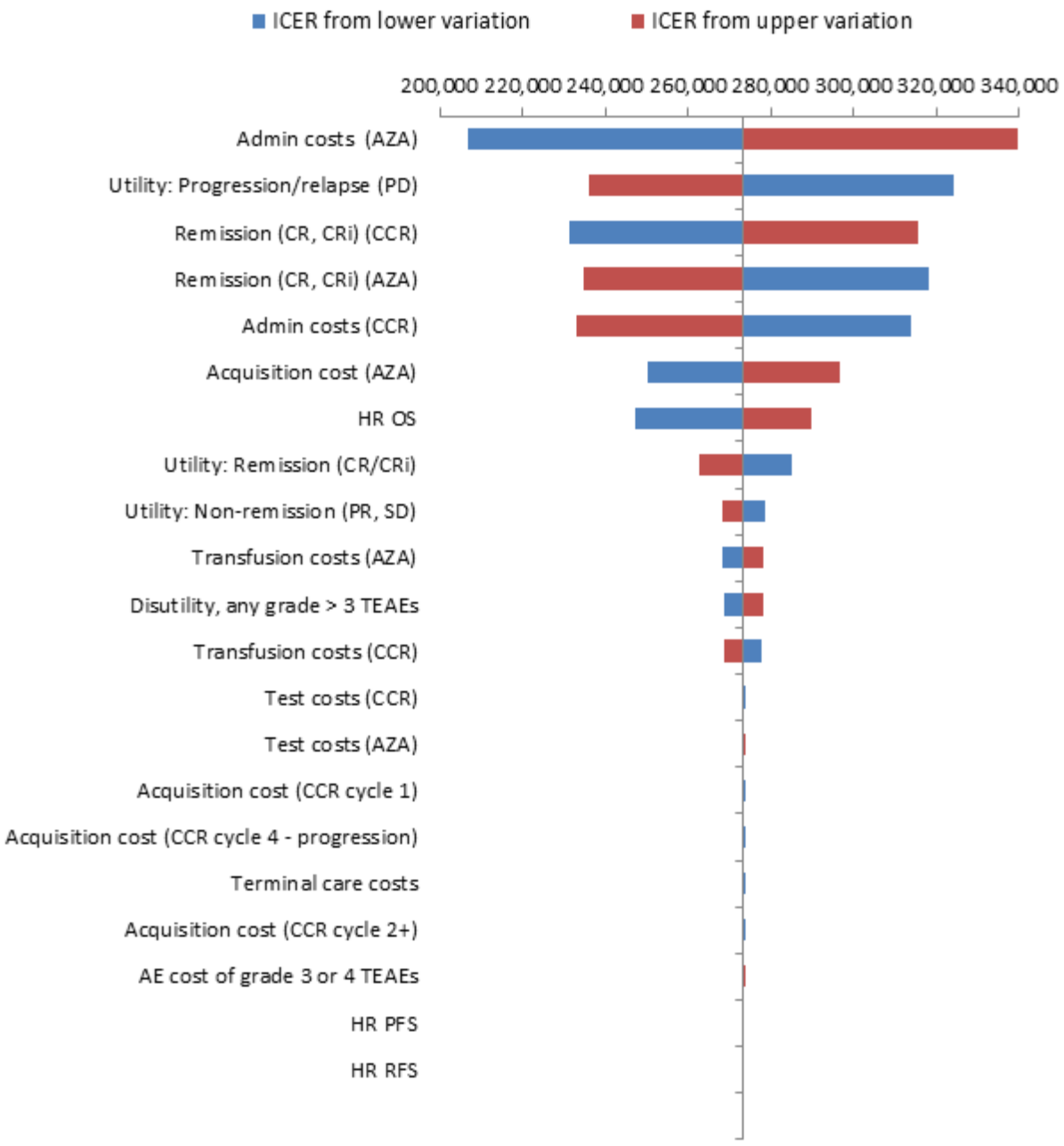
Notes: a, Negative ICERs indicate azacitidine is dominant; b, Includes corrections and changes as described in *Table 59* except for component 'G' (i.e., not including adjustment for baseline covariates); c, Not including component 'C' (i.e., retaining Celgene's estimates for costs in Relapse/PD)

To assess the validity of these analyses, the ERG derived an ICER for the whole population using a weighted average of the incremental costs and QALYs across the three CCR therapy preselected subgroups. The resulting ICER was £269,714 per QALY, compared to an ICER of £246,488 per QALY using changes A–F for the whole population (*Table 59*). This is a discrepancy of less than 10%.

6.2 Univariate sensitivity analyses

The univariate sensitivity analysis with the base case preferred by ERG is presented in the tornado analysis of *Figure 23*; plausible variation of parameter values results in ICERs above £200,000 per QALY.

Figure 23: Tornado diagram of ERG's preferred base case deterministic analysis



Key: AZA, azacitidine; 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; RFS, relapse-free survival; SD, stable disease; TEAE, treatment emergent adverse event

6.3 Probabilistic sensitivity analysis

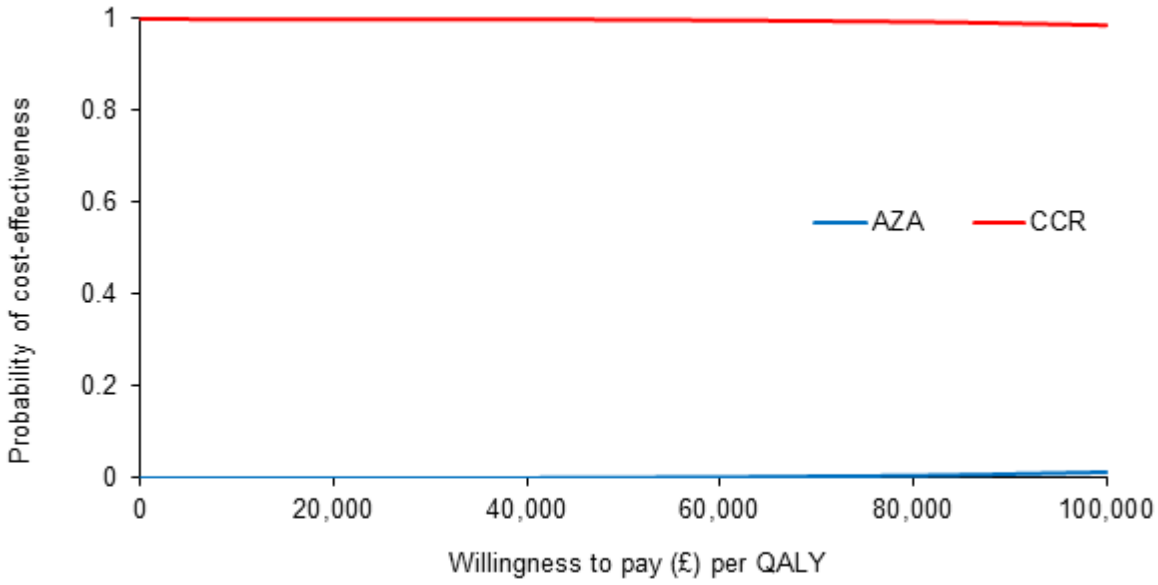
A probabilistic sensitivity analysis was conducted for the ERG's preferred base case (*Table 61*). An ICER of £277,123 per QALY was obtained, which is similar to the deterministic ICER of £273,308 per QALY. *Figure 24* presents the cost-effectiveness acceptability curves from the probabilistic sensitivity analysis in the ERG's preferred base case. At a willingness to pay threshold of £100,000 the probability of azacitidine being cost-effective is less than 5%.

Table 61: Cost-effectiveness results for ERG's preferred base case probabilistic sensitivity analysis

Arm	Total			Incremental			ICER (cost per QALY)
	Costs	LYG	QALYs	Costs	LYG	QALYs	
CCR	£73,152	0.8863	0.6218	—	—	—	—
Azacitidine	[redacted]	1.3302	[redacted]	[redacted]	0.4439	[redacted]	£277,123

Key: CCR; conventional care regimens; ERG, Evidence Review Group; ICER, incremental cost-effectiveness ratio; LYG, life years gained; QALYs, quality-adjusted life years

Figure 24: Cost-effectiveness acceptability curves from ERG's preferred base case probabilistic sensitivity analysis



Key: AZA, azacitidine; CCR, conventional care regimens; ERG, Evidence Review Group; QALY, quality-adjusted life year

Azacitidine for treating acute myeloid leukaemia with more than 30% bone marrow blasts

Errata

Location in report	Original text	Corrected text
Section 1.7, p.17	The ERG preferred base case ICER is £169,606 compared to the company's base case ICER of £20,648 per QALY.	The ERG preferred base case ICER is £273,308 compared to the company's base case ICER of £20,648 per QALY.
Section 1.7, p. 17	Adjusting overall survival for baseline covariates (reduces ICER from £246,488 to £169,606 per QALY).	Adjusting overall survival for baseline covariates (reduces ICER from £246,488 to £273,308 per QALY).
Section 5.3.1, p. 99	EQ-5D survey mapped from disease specific single trial data using a published mapping algorithm.	EQ-5D survey mapped from disease specific single trial data using a published mapping algorithm; a second published algorithm was used for sensitivity analysis.
Section 5.3.1, p. 102	Cycle length is 4 months.	Cycle length is 4 weeks.
Section 5.3.1, p. 103	An error was found in the way parameter values for the number of treatment cycles was inputted in the model, which resulted in a large underestimation in costs of drug acquisition, monitoring tests, transfusions and the company's drug ICER	An error was found in the way parameter values for the number of treatment cycles was inputted in the model, which resulted in a large underestimation in costs of drug acquisition and administration, monitoring tests, transfusions and the company's drug ICER.
Section 5.3.4, p. 106	However, the way parameter values for the number of treatment cycles were implemented in the model was incorrect, resulting in a mean number of treatment cycles in the azacitidine group of 5.6 instead of the intended 8.8, in the CCR group IC of 1.86 instead of 2 (initiation and consolidation) and 4.4 when estimating drug acquisition costs and 5.3 when calculating the costs of drug administration, tests and transfusion instead of 6.10 in the CCR group LDAC.	However, the way parameter values for the number of treatment cycles were implemented in the model was incorrect, resulting in a mean number of treatment cycles in the azacitidine group of 5.6 instead of the intended 8.8, in the CCR group IC of 2.61 instead of 2 (initiation and consolidation), 4.4 when estimating drug acquisition costs and 5.3 when calculating the costs of drug administration, tests and transfusion instead of 6.10 in the CCR group LDAC.

Section 5.3.6.2, p114	The ERG undertook diagnostic tests of the proportional hazards assumption on the adjusted Cox PH OS curves by comparing the cumulative log-log plot of the azacitidine and CCR arms directly as well as testing for time and treatment effect interactions and Schoenfeld residuals (see details in *Appendix 2*). The results of these tests suggest the constant proportional hazards assumption is not supported by the data.	The ERG undertook diagnostic tests of the proportional hazards assumption on the adjusted Cox PH OS curves by comparing the cumulative log-log plot of the azacitidine and CCR arms directly by Schoenfeld residual test; the proportional hazard assumption is not rejected at the 5% level, p=0.068, but graphical inspection suggest the test may not be robust and cast doubt on the appropriateness of the assumption in these data (see details in *Appendix 2*).
Section 5.3.6.2, p.117, Table 53	Bathtub Adjusted for baseline covariates, Azacitidine, AIC 640, BIC	Bathtub Adjusted for baseline covariates, Azacitidine, AIC 648, BIC 718
Section 5.3.6.2, p.117, Table 53	Bathtub Adjusted for baseline covariates, CCR, AIC 645, BIC	Bathtub Adjusted for baseline covariates, CCR, AIC 646, BIC 716
Section 5.3.6.2, p.117, Table 53	Bathtub Notes	Bathtub[c] Notes: c, Bathtub survival function $\exp(-(\exp[(\lambda t)^{\rho}]-1))$. Further details available from ERG upon request.
Section 5.3.6.2, p.117	The adjusted exponential model fitted to both arms results in a HR of 0.64 and has a predicted difference in OS of 3.64 months, in favour of azacitidine; details are presented in the *Appendix 2*.	The adjusted exponential model fitted to both arms results in a HR of 0.65 and has a predicted difference in OS of 3.64 months, in favour of azacitidine; details are presented in the *Appendix 2*.
Section 5.3.6.2, p. 117	*Figure 22* presents the ITT Kaplan-Meier data and the fitted adjusted exponential OS model to data censored at switch to subsequent AML therapy in both trial arms.	*Figure 22* presents the Kaplan-Meier data (censored at switch) and the fitted adjusted exponential OS model to data censored at switch to subsequent AML therapy in both trial arms.
Figure 22, p. 118 (Caption)	Figure 22: Overall survival in AZA-AML-001 – ITT Kaplan-Meier data and adjusted exponential model fitted to censor-at-switch (any AML therapy) data	Figure 22: Overall survival in AZA-AML-001 – Kaplan-Meier data (censored at switch) and adjusted exponential model fitted to censor-at-switch (any AML therapy) data
Figure 22, p. 118	[Figure]	[See Figure 22 below]
Section 5.3.6.2, p.119	As a conservative approach, the ERG's preferred base case analysis adopted the exponential OS HR estimates adjusted for baseline covariates used by the IPCW method, including sex, age, ECOG status, preselected CCR treatment, time since initial AML diagnosis, comorbidity score, in data censored at switch to any subsequent AML treatment from both trial arms in the dataset provided by Celgene to ERG. The resulting HR estimate, 0.64, is more favourable to azacitidine than the respective estimates from applying IPCW to subsequent azacitidine use in CCR	As a conservative approach, the ERG's preferred base case analysis adopted the exponential OS HR estimates adjusted for baseline covariates used by the IPCW method, including sex, age, ECOG status, preselected CCR treatment, time since initial AML diagnosis, comorbidity score, in data censored at switch to any subsequent AML treatment from both trial arms in the dataset provided by Celgene to ERG. The resulting HR estimate, 0.65, is more favourable to azacitidine than the respective estimates from applying IPCW to subsequent azacitidine use in CCR

	only, [redacted] and [redacted] (Celgene's base case), IPCW to any AML in both arms arms, 0.77 and 0.71, and the ITT value of 0.85.[45]	only, [redacted] and [redacted] (Celgene's base case), IPCW to any AML in both arms arms, 0.77 and 0.71, and the ITT value of 0.85.[45]
Section 5.3.6.2, p.120	The subgroup analyses were thus based on the original censor-at switch exponential OS curves estimates used by the company in their model.	The subgroup analyses were thus based on the original censor-at switch exponential OS curves estimates used by the company in their model (alongside Kaplan-Meier curves for censor at switch PFS and RFS data available in the Celgene excel model).
Section 5.3.7, p.120	The QALY impact of AEs was modelled as the probability of at least one TEAE of Grade > 3 per 100 person-years multiplied by the utility of grade > 3 TEAEs. In the model, the effect of AEs on QALYs in azacitidine have been modelled for a maximum of 8.8 treatment cycles, which is effectively less than the mean 8.8 in the trial, while for CCR as a whole they were counted for a maximum number of cycles of 5.1 which is more than the mean number of 2 cycles with IC. These calculations are likely to overestimate the additional costs of CCR relative to azacitidine but this bias had a small effect on the results given other issues identified by the ERG in the model.	The QALY impact of AEs was modelled as the probability of at least one TEAE of Grade ≥ 3 per 100 person-years multiplied by the utility of grade ≥ 3 TEAEs. In the model, the effect of AEs on QALYs in azacitidine have been modelled for a maximum of 8.8 treatment cycles, which is effectively a mean of 6.43 in the model rather than the mean 8.8 cycles in the trial, while for CCR as a whole they were counted for a maximum number of cycles of 5.1, which resulted in a mean of 4.23 treatment cycles instead of the mean number of 2 cycles with IC in the trial. These calculations are likely to overestimate the additional QALY losses due to disutility of CCR relative to azacitidine treatment but this bias had a small effect on the results given other issues identified by the ERG in the model.
Section 5.3.8.1, p. 122	Likewise the correct maximum number of cycles for LDAC in CCR is 10 cycles (the 1.6 cycles used for IC in CCR modelled by Celgene is practically correct). In correcting the probabilistic sensitivity analysis ERG chose SE 3 for azacitidine, 1 LDAC and (0.06 for IC as used by Celgene's base case is correct), to calibrate the respective model outputs to the reported figures for mean minus one standard error in the number of cycles reported by Dombret et al. 2015.[45] As documented below, correcting this error of implementation to calibrate the model outputs with the summary statistic reported by the trial increases the ICER from a level around £20,000 to £84,000 per QALY gained.	Likewise the correct maximum number of cycles for LDAC in CCR is 10 cycles and 2 cycles for IC in CCR. In correcting the probabilistic sensitivity analysis ERG chose SE 3 for azacitidine, and 1 for LDAC (0.06 for IC as used by Celgene's base case is correct), to calibrate the respective model outputs to the reported figures for mean minus one standard error in the number of cycles reported by Dombret et al. 2015.[45] As documented below, correcting this error of implementation to calibrate the model outputs with the summary statistic reported by the trial increases the ICER from a level around £60,000 (Celgene's corrected base case) to around £130,000 per QALY gained.
Section 5.3.8.2, p.123	Another limitation was that the costs of managing AEs in patients from azacitidine and CCR groups were estimated as the product of the probability of at least one TEAE of	Another limitation was that the costs of managing AEs in patients from azacitidine and CCR groups were estimated as the product of the probability of at least one TEAE of

	Grade > 3 per 100 person-years and the average cost of managing grade 3 or 4 TEAEs observed in the AZA-AML-001 trial.	Grade ≥ 3 per 100 person-years and the average cost of managing grade 3 or 4 TEAEs observed in the AZA-AML-001 trial.
Section 5.3.10, p. 126	Our corrected results presented in *Section 6* (*page 130*) are subject to this caveat.	However, the ERG corrected base case results presented in *Section 6* (*page130*) were found to be replicable by the weighted average method within a 5-10% margin of error.
Section 5.4.2.3, p.128	The ERG set the OS curves to the Censor At Switch analysis mode in Celgene's model, keeping the exponential functional form adopted in Celgene's base-case. This choice was associated with a hazard ratio of 0.72, which corresponded to the analysis that was unadjusted for baseline covariates. Given the high ICERs that were obtained after the preceding revisions in *Sections 5.4.2.1* and *5.4.2.2*, the effect of adopting the smaller hazard that resulted from censored at switch analysis that adjusted for baseline covariates was investigated in exploratory analyses.	The ERG set the OS curves to the Censor At Switch analysis mode in Celgene's model, keeping the exponential functional form adopted in Celgene's base-case. This choice was associated with a hazard ratio of 0.72, which corresponded to the analysis that was unadjusted for baseline covariates. In addition, the Censor At Switch analysis in the Celgene model alters the estimates of relative effects of azacitidine on PFS and RFS, from the Celgene base case HR values of 0.84 and 0.85 to 0.83 and 0.76, respectively. Given the high ICERs that were obtained after the preceding revisions in *Sections 5.4.2.1* and *5.4.2.2*, the effect of adopting the smaller hazard that resulted from censored at switch analysis that adjusted for baseline covariates was investigated in separate exploratory analyses.
Secton 5.4.2.5, p. 128	The effect of using the OS hazard ratio estimate of 0.64 from the exponential model adjusted for baseline covariates, which found support in statistical tests conducted by the ERG, was investigated.	The effect of using the OS hazard ratio estimate of 0.65 from the exponential model adjusted for baseline covariates, which found support in statistical tests conducted by the ERG, was investigated.
Section 5.4.3, p. 128	The ERG sought to perform some exploratory assessment of the subgroup analysis by preselected CCR treatment, while acknowledging that for PFS and RFS outcomes, the sample sizes make subgroup-specific time to event data highly unreliable. Thus in these analyses subgroup specific differences in OS outcomes were allowed using censor-at-switch data, while keeping common PFS and RFS curves across the three subgroups.	Since resource use data in the Celgene model was obtained from a survey of clinical experts' estimates, and because the ERG's preferred cost estimates for Relapse/PD were those estimated by Celgene's experts for the BSC (i.e. the most costly) subgroup during Relapse/PD, the effect of uncertainty from these values on the ICER was explored in extreme scenarios where costs of Relapse/PD were zero and, separately, where no inpatient costs were incurred (apart from what is already included in the model 'Terminal care' costs). In its preferred base case analysis, Celgene measured costs of monitoring tests and transfusions only while patients were on azacitidine, LDAC or

		IC. The ERG instead included costs for these items for the duration of the patients in Remission and Non-Remission states as well as during Relapse/ Progressive Disease. The ERG explored the effects of adopting Celgene's assumption that no such costs would occur after treatment active treatment stopped. Other scenarios included revising the costs of monitoring tests and transfusions during Relapse/PD to values estimated by a clinical expert consulted by the ERG, whereby two units each of red blood cell transfusions and adult doses of platelet transfusions are given on average per 4 week cycle, whilst no bone marrow aspirates or biopsies nor extractions for cytogenetic testing, four blood tests, and two each of peripheral blood smears and serum blood chemistry are given during PD. In addition, the ERG sought to perform some exploratory assessment of the subgroup analysis by preselected CCR treatment, while acknowledging that for PFS and RFS outcomes, the sample sizes make subgroup-specific time to event data highly unreliable. In these analyses subgroup specific differences in OS, PFS, and RFS outcomes were allowed using censor-at-switch data, using the exponential (OS) and Kaplan-Meier (PFS and RFS) curves provided by Celgene in the Excel model for each arm in the three subgroups.
Section 6, pp. 130–136	[Section 6 and all included tables (58 to 61) and figures (23 and 24) superseded by the ERG Addendum following correction of an error in the final stage of the derivation of the ERG preferred base case]	
Section 8, p. 138	Of the key changes made to the economic model by the ERG in arriving at their preferred base case (ICER £169,606 per QALY), the following might be considered to be differences of opinion between the company and the ERG:	Of the key changes made to the economic model by the ERG in arriving at their preferred base case (ICER £273,308 per QALY), the following might be considered to be differences of opinion between the company and the ERG:
Section 8, p. 138	The base case analysis preferred by the ERG has the limitation that the sum of the cohort proportions in PFS and RFS exceeds the proportion alive by the third cycle in the model. Celgene included an adjustment in the model to eliminate such anomalies, but nevertheless this remains a deficiency of the model structure. The ERG has not attempted to correct this, due to the	[This is no longer the case after correcting the error in estimating the ERG's adjusted exponential OS baseline hazard; Appendix has been corrected below]

	difficulties of fitting plausible parametric models to the PFS and RFS data (as discussed in this critique) and adjusting for baseline covariates, given the small samples available for analysis.	
Appendix 2, Figure A1 (caption)	Figure A1: Log-log plots of OS probability adjusted for subsequent azacitidine use in the CCR arm using the IPCW method – unadjusted for baseline covariates	Figure A1: Log-log plots of OS probability adjusted for subsequent azacitidine use in the CCR arm using the IPCW method – adjusted for baseline covariates
Appendix 2, Figure A1, footnotes	Note: Not adjusted for differences in baseline covariates	[Note deleted]
Appendix 2, Figure A1	[Figure]	[See Figure A1 below]
Appendix 2, Section A2.1		Append following subsection: A2.1.2 Schoenfeld residual plots Figure A2: Schoenfeld residual plots and local polynomial regression, adjusted for subsequent azacitidine use in the CCR arm – adjusted for baseline covariates [Figure A2] Key: CCR, conventional care regimens Source: ERG analysis using individual patient data provided by Celgene
Appendices 2–4	[Figure numbering]	[Renumber figures]
Appendix 2, Section A2.4	Model estimated using effects coding to obtain average baseline. Mean time to death Azacitidine = 0.0973354^{-1} = 10.27 months; CCR =$(1.549388 \times 0.0973354)^{-1}$ = 6.63 months (HR azacitidine vs. CCR = 1.549388^{-1} = 0.6454162)	Model estimated using effects coding to obtain average baseline. Mean time to death Azacitidine = 0.058019^{-1} = 17.24 months; CCR =$(1.549388 \times 0.058019)^{-1}$ = 11.12 months (HR azacitidine vs. CCR = 1.549388^{-1} = 0.6454162)
Appendix 2, Section A2.4	This model adjusted for baseline covariates sex, age, ECOG, cytogenetic risk, CCR preselected treatment, comorbidity group, AML days, platelet transfusion status, geographical region, used in Celgene's IPCW analysis (fixed covariates), using effects-coding, in order to estimate the average baseline (intercept) coefficient in the sample. Blast group was not effects-coded but included in order to adjust for blast group <25 (blastgrp2 indicates 5-24% blasts; blastgrp4 indicates blastgrp<5%), so that estimated baseline coefficient of 0.0973354 is average in sample with blasts≥25% group.	This model adjusted for baseline covariates sex, age, ECOG, cytogenetic risk, CCR preselected treatment, comorbidity group, AML days, platelet transfusion status, geographical region, used in Celgene's IPCW analysis (fixed covariates). The predicted baseline hazard rate was obtained by evaluating the estimated equation and mean valued of the covariates for the whole. Blast group was not evaluated at 0 in order to adjust for blast group <25 (blastgrp2 indicates 5-24% blasts; blastgrp4 indicates blastgrp<5%), so that estimated baseline coefficient of 0.058019 is average in sample with blasts≥25% group.

Appendix 2, Section A2.4	[Stata output from streg command]	[Stata output as shown below]

Figure 22: Overall survival in AZA-AML-001 – Kaplan-Meier data (censored at switch) and adjusted exponential model fitted to censor-at-switch (any AML therapy) data

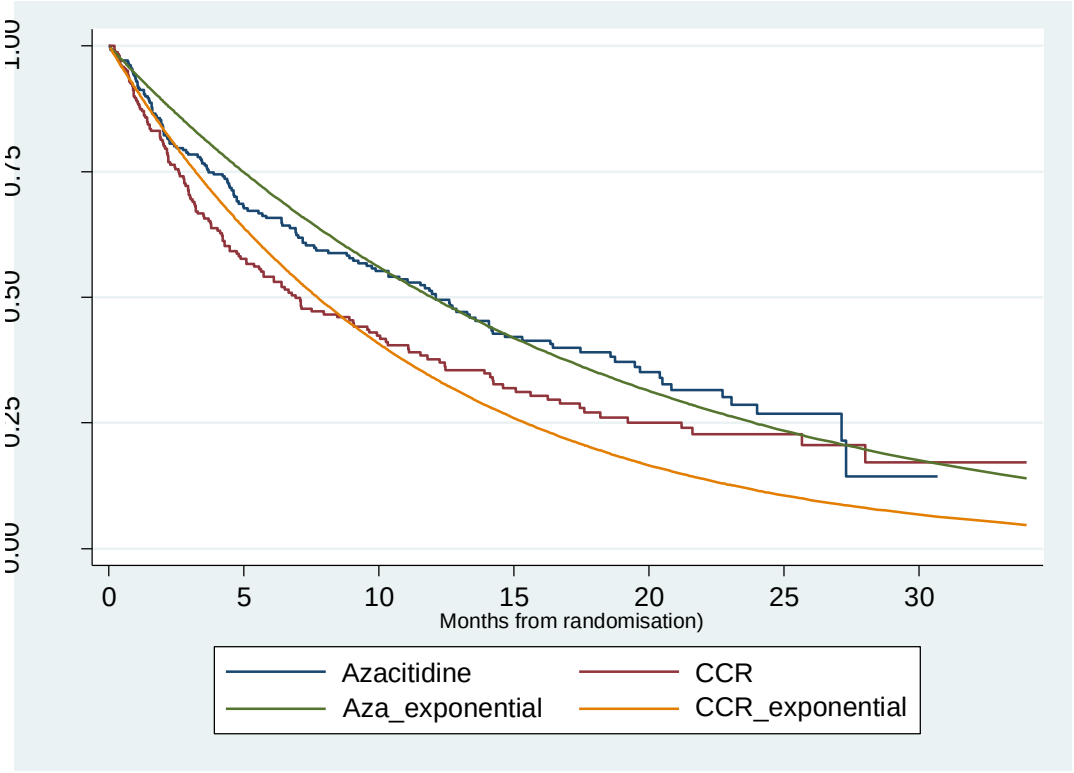


Figure A1: Log-log plots of OS probability adjusted for subsequent azacitidine use in the CCR arm using the IPCW method – adjusted for baseline covariates

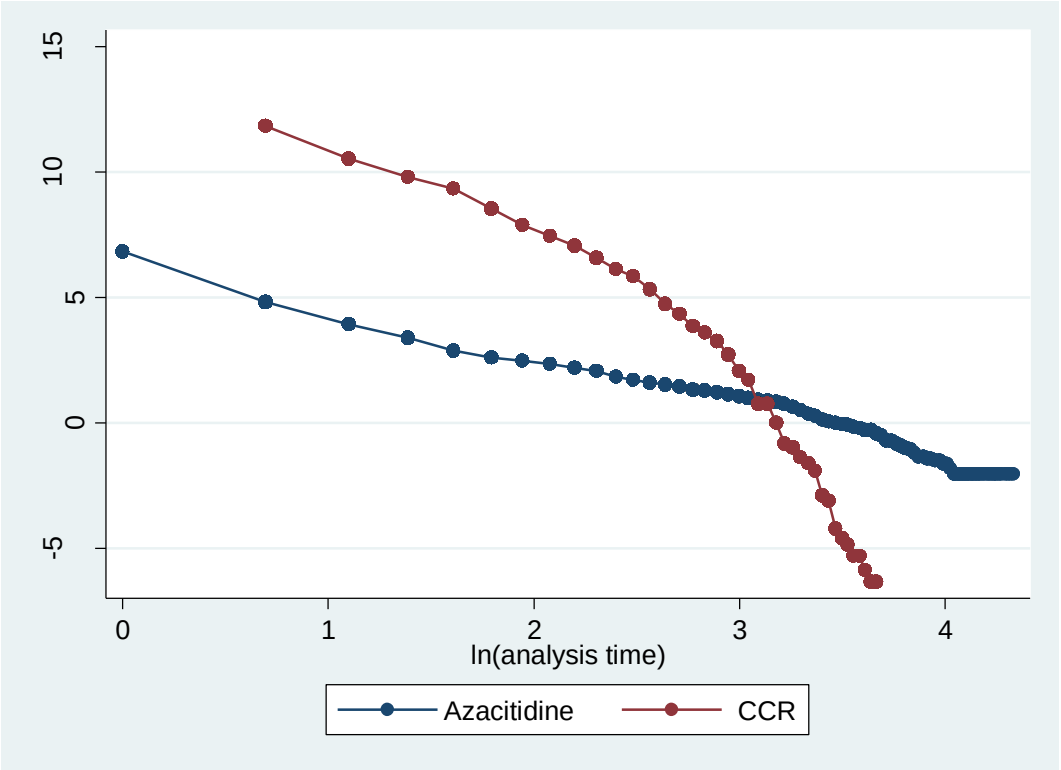
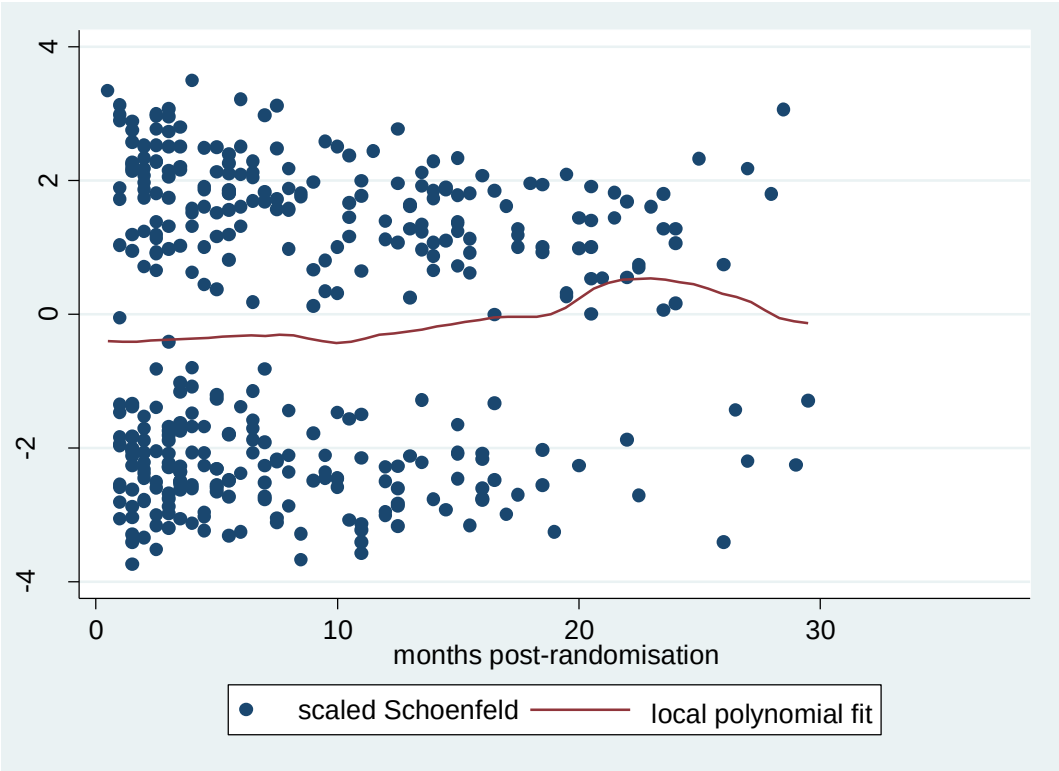


Figure A2: Schoenfeld residuals and local polynomial regression, adjusted for subsequent azacitidine use in the CCR arm – adjusted for baseline covariates



Revised Stata output for streg command (Section A2.4)

_t	Haz. Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
ccr	1.549388	.1874601	3.62	0.000	1.222288	1.964024
sex2	1.028439	.126573	0.23	0.820	.8080132	1.308996
agegrp2	1.633048	.2117287	3.78	0.000	1.266597	2.10552
ECOG2	1.490963	.2070571	2.88	0.004	1.135681	1.957389
CYTO2	2.117182	.2584642	6.14	0.000	1.666649	2.689505
blastgrp2	.379288	.149469	-2.46	0.014	.1751981	.8211239
blastgrp4	.1908186	.1924331	-1.64	0.100	.0264373	1.377286
randpr2	.3804746	.0875254	-4.20	0.000	.242389	.5972257
randpr3	.4458237	.0697034	-5.17	0.000	.3281557	.6056844
comorbis3	1.386397	.220576	2.05	0.040	1.014989	1.893712
aml1	1.033729	.3199961	0.11	0.915	.5635249	1.896272
aml2	1.181031	.358463	0.55	0.584	.6514894	2.140994
aml3	.933583	.2925007	-0.22	0.826	.5051998	1.725213
pltstat2	.6219478	.0802252	-3.68	0.000	.4830113	.8008489
georeg1	1.159714	.2316993	0.74	0.458	.7839519	1.715586
georeg2	2.366019	.4000999	5.09	0.000	1.698549	3.295783
georeg3	1.194632	.2019344	1.05	0.293	.8577316	1.66386
_cons	.0591534	.0223448	-7.49	0.000	.0282126	.1240269

National Institute for Health and Care Excellence
Centre for Health Technology Evaluation

Pro-forma Response

ERG report

Azacitidine for treating acute myeloid leukaemia with more than 30% bone marrow blasts [ID829]

You are asked to check the ERG report from Peninsula Technology Assessment Group (PenTAG) to ensure there are no factual inaccuracies contained within it.

If you do identify any factual inaccuracies you must inform NICE by **5pm, Friday 19 February 2016 (changed to Wednesday 24th February 2016)** using the below proforma comments table. All factual errors will be highlighted in a report and presented to the Appraisal Committee and will subsequently be published on the NICE website with the Evaluation report.

The proforma document should act as a method of detailing any inaccuracies found and how and why they should be corrected.

Issue 1 Use of CCR as a comparator; not investigating individual treatments

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
Throughout the report (example Table 1, pg. 24), the ERG have	It should be stated that:	Stating that the company have not assessed whether azacitidine	The company have not *adequately* assessed whether azacitidine

stated that CCR was used as a comparator rather than IC LDAC and BSC, and as a result the company have not assessed whether azacitidine demonstrates clinical and cost effectiveness compared with each treatment within CCR	The company have used CCR in their base case analysis; however they have provided sensitivity analyses comparing patients receiving each of the treatments making up CCR	demonstrates clinical and cost effectiveness compared with each treatment within CCR is factually inaccurate, and leads the reader to conclude that the company have not investigated individual treatments.	demonstrates clinical and cost-effectiveness compared with each treatment within CCR for the following reasons: 1. The pivotal RCT was not powered for such comparisons. 2. There are significant inconsistencies between the results of the company's base case and subgroup analyses (ERG report, p. 125). In the stated example the ERG considers it is clear that the focus of the critique is the base case. The ERG report does include the subgroup analyses and therefore a full reading of the ERG report would not lead the reader to conclude that the company have not investigated individual treatments. **No action taken.**

Issue 2 Suggestion that the literature review for clinical effectiveness studies was poor

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
On pg. 28, the ERG state "The ERG believes the literature searching for clinical effectiveness studies was poorly conducted and reported". They also state that their own	Suggested text: The methodology used in the literature, although less thorough than expected, did not miss any	It is factually inaccurate to suggest that the methodology was poor if it identified all relevant studies.	The ERG rejects this proposed amendment. The ERG critique remains an accurate reflection on the quality of the company's submission

searches did not identify any additional studies.	studies found by the ERG's own literature search.		as it relates to study identification. **No action taken.**

Issue 3 The dropout rates due to an AE look imbalanced for azacitidine (36.9%) compared to the CCR group (26.7%)

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
Table 3 pg.36 the ERG state "The drop-outs provided by Celgene represent the figures reported in the RCT. However, neither Celgene nor the RCT report the drop outs for the three CCR treatments separately, therefore it is unknown whether the actual treatments that make up the CCR are comparatively different to azacitidine. The dropout rates due to an AE look imbalanced for azacitidine (36.9%) compared to the CCR group (26.7%)". This is a misleading statement.	Text should state: Within the single technology appraisal (STA) company evidence submission document, dated 25 November 2015, Figure 7, the CONSORT diagram for AZA-AML-001 depicts the disposition of subjects in the intent-to-treat (ITT) population for the Treatment Phase, with each of the three CCR treatment arms presented individually. Also, as seen in the STA submission document, the duration of therapy is considerably longer in the azacitidine group compared with the individual CCR groups and this must be considered when interpreting these data. In Study AZA-AML-001, the median duration of treatment for the azacitidine group (164.5 days) was longer than that observed for the BSC-only (65.0 days), low-dose cytarabine (98.0 days), or intensive chemotherapy (55.5 days) groups.	The original statement in the ERG report is misleading without providing text to explain this difference.	The company is correct to indicate that Figure 7 provides the detail which the ERG originally said was not reported. **The ERG agrees to change the statement in Table 3 to:** The ERG has no further comments.

	The total treatment exposure expressed in person-years was higher in the azacitidine group (174.9) than in the other groups (82.9, 14.1, and 9.6 for low-dose cytarabine, intensive chemotherapy, and BSC-only groups, respectively).		

Issue 4 Statement that none of the published registry data provided information on the population with >30% blasts

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
On page 66 of the ERG report, it is stated that "None of the published registry data (Austrian azacitidine registry (NCT01595295); Spanish AMLA registry and French compassionate patient named programme) provided information on the population with >30% bone marrow blasts, a population requirements for the Scope for this report."	This statement should be removed.	The registry data discussed do include information on patients with >30% bone marrow blasts; therefore the ERG's statement is factually inaccurate.	The ERG's intended meaning was that results pertaining only to the subpopulation with >30% blasts was not presented (not that patients with >30% blasts were not represented at all). However, the ERG acknowledges that results were presented for a relevant subpopulation from the Austrian Azacitidine Registry. **The ERG agrees to replace the paragraph with the following:** Both the Spanish AMLA registry and French compassionate patient named programme failed to provide outcomes for the specific population with >30% bone marrow blasts. Instead only the proportion of patients with >30% bone marrow

			blasts was reported. Outcomes were reported combined irrespective of bone marrow blast status. The Austrian azacitidine registry (AAR, NCT01595295) report outcomes for the total patient population irrespective of bone marrow blasts status. In addition, they also report outcomes for those with >30% bone marrow blasts and a WBC <15g/L receiving azacitidine and compare baseline and treatment characteristics for these patients to the azacitidine arm of the AZA-AML-001 trial. Baseline and treatment characteristics were similar between the two groups of patients receiving azacitidine except for the following differences: Baseline characteristics for AML classification: AML-NOS was higher (63.5 %) in the AZA-AML-001 trial than the AAR (24.2%) and for AML-MRF, the AAR reported higher proportions (66.3%) than the AZA-AML-001 trial (31.1%). Outcomes appear to be similar for the registry and the AZA-AML-001 trial. For AE, treatment-emergent thrombocytopenia and anaemia both had significantly higher (47.4% and 31.6% respectively) incidences in the

			Austrian azacitidine registry than from the AZA-AML-001 trial (15.7% and 26.3% respectively).

Issue 5 Statement that the ERG's preferred base case equalises the costs in the relapse/progressive disease health state across the model

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
Throughout the report, the ERG make reference to their preferred base case of setting relapse/PD health state costs to be equal across the model. For example on pg.15 they say "The ERG also identified that there were significant differences in the cost associated with the Relapse/progressive disease state between the AZA and CCR arm, even though all patients (in both arms) are expected to be receiving BSC at this point." This is further detailed in section 5.3.8.2 on pages 122-123. In Appendix 3 of the ERG report, the ERG state that they have applied the data for the BSC cohort to all cohorts within the relapse/PD state. The ERG also state that it is questionable that the sequence of treatments studied in the model i.e., AZA followed by BSC and CCR	A statement should be added by the ERG in relation to their preferred base case, that setting setting relapse/PD health state costs to be equal across the model is not reflective of the AZA-AML-001 trial, the conclusions drawn by the clinical questionnaire designed to elicit resource use (from 7 UK clinicians), or the clinical opinion given to the ERG. In addition, it should be noted that applying the BSC cohort data to all cohorts in the relapse/PD state is inappropriate and not representative of clinical practice.	There is inconsistent messaging from the ERG; one message of setting relapse/PD health state costs to be equal across the model, and the other than patients would receive active treatment after 1st line. The combination of the AZA-AML-001 trial, the conclusions drawn by the clinical questionnaire designed to elicit resource use (from 7 UK clinicians), as well as clinical opinion given to the ERG, suggests that patient resource use is likely to vary after 1st line treatment. Applying the BSC cohort data to all cohorts in the relapse/PD state is inappropriate and not representative of clinical practice.	The proposed statements are not appropriate. The ERG had to work within the limitations of the structure of the Celgene model. The ERG's limited aim was to correct the OS effectiveness parameter values in Celgene's model to make them consistent with the model's assumption that no subsequent active therapy would be used in AZA or CCR after 1st line treatment; see Celgene submission, Section 5.6.2., Table 54, p. 141 "Assumption: There is no treatment switching. Justification Clinical expert opinion. Only a very small percentage of patients at this stage of disease would be fit for a second treatment after failing their first." ERG's corrections do not imply endorsement of Celgene's assumptions. The model should have

followed by BSC) is realistic, suggesting subsequent therapy may be given in clinical practice (as was seen in the AZA-AML-001 trial). In addition, the ERG states (pg. 129) "Advice from clinical experts suggests that active second-line treatment is considered for some patients in the NHS".			allowed for the cost and effects of subsequent treatments used in the trial which would also be used in routine NHS practice, by modelling second line treatment. However, Celgene's model assumed that no such treatment was available after relapse/PD. As stated in ERG's report, Section 5.3.6.2, p. 116. "To remedy the contradiction between the model structure, on the one hand, and the methodology underlying the OS treatment effect estimates, on the other, two options were available to the ERG. One was to correct the model to include the costs of subsequent treatments used in the AZA-AML-001 and left unadjusted for in the statistical analysis that produced the base case OS treatment effect estimates. This option was not feasible because the required data on the dates of start and end of subsequent treatments as well as treatment dosages and frequencies of administration were not available to ERG." Note that setting relapse/PD costs to be equal across the model was simply done for consistency with the model assumptions; i.e. Celgene assume that patients in this state are all managed by BSC and it was not clear why there should be differences in costs by initial treatment (i.e. AZA

			vs CCR). Different costs may apply to the relapse/PD phase (and ERG presented results of varying these), but claiming that costs should be different between cohorts is questionable given the lack of data and difficulties of defining these CCR preselected subgroups a priori. Further, Celgene did not provide any evidence in its submission that the resource use survey addressed the role of the specific subsequent treatments used in AZA-AML-001. **No action taken.**

Issue 6 Statement that combined PFS and RFS curves cross OS curve at cycle 69

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
On page 110, the ERG state "The inadequacy of the statistical analysis may be illustrated by comparing the estimated time to event curves used by Celgene in their base case analysis, combining PFS with RFS and contrasting it with OS. **Error! Reference source not found.** below illustrates the problem for the CCR arm. By cycle 69, that is, just after the start of the sixth year after first receiving treatment the estimated curves imply that there are more	This statement should be removed.	This is factually inaccurate. Having checked the model, the curves do not cross. A possible reason for the ERG's interpretation may be that they have compared the wrong settings: RFS and PFS in columns H and I of the Model worksheets can exceed OS in column J because they are comprised from different denominators: in fact, OS from column J must be compared with	The ERG stands by its comment. The ERG confirms it has considered the appropriate columns. The reason these curves, as implemented in the Excel model, do not cross is that in the Excel model Celgene applied an adjustment so that when crossing would have occurred the survival probabilities distributed the OS time proportionately to PFS and RFS; e.g. in Celgene Model AZA sheet cell L23 the formula is:

patients in the treated cohort who are either in Remission or non-remission (stable disease) than there are patients alive.”		RFS and PFS from columns K and L, respectively.	= IF((RFS*'PSA inputs'!$D$22+PFS*'PSA inputs'!$D$24)>OS,OS*'PSA inputs'!$D$22/('PSA inputs'!$D$22+'PSA inputs'!$D$24),RFS*'PSA inputs'!$D$22). This correction avoids having PFS + RFS larger than OS, but does not address concerns about the validity of Celgene’s chosen parametric survival curves for assigning time to Remission and Non-remission versus relapse/PD states in the model. **No action taken.**

Issue 7 The use of censor at switch analyses

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
On pages 116 to 120, the ERG describe using censor at switch methodology to model the OS. They acknowledge that they did this because they were not able to perform the IPCW analysis to adjust for subsequent therapy for both treatment groups. The ERG also acknowledge that (pg. 129) “Advice from clinical experts	It should be noted along with the ERG’s preferred base case ICER that the methodology used for modelling OS is not optimal, and does not fit with the AZA-AML-001 trial, the conclusions drawn by the clinical questionnaire designed to elicit resource use (from 7 UK clinicians), as well as clinical opinion given to the ERG, all of which suggest patients could receive	The ERG’s base case ICER is misleading without acknowledging the limitations of the OS modelling whenever the base case is mentioned.	The ERG’s preferred base case is the best possible use of the available data given Celgene’s model logic. It does not represent ERG’s view of what UK routine practice is or an endorsement of Celgene’s structural assumptions, whose correction was beyond the scope of ERG’s STA review as defined by NICE. We acknowledge this in ERG report Section 5.3.6.2, p. 117: “Whether the

suggests that active second-line treatment is considered for some patients in the NHS"	subsequent therapy.		assumption that no subsequent active treatment would be available to UK patients in routine practice is plausible may of course be questioned, but the point here is that applying IPCW adjustment for subsequent treatment to the outcome data of the CCR but not the azacitidine arm in the AZA-AML-001 trial is inconsistent with the economic model which the overall survival IPCW analysis was designed to inform." Please also see our response to Issue 6. **<u>No action taken.</u>**

Issue 8 OS modelling methods and results are unrealistic and misleading

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
The ERG describes their preferred base case for OS modelling (pg. 116-120), as well as the use of KM data for PFS and RFS (pg. 128). The ERG state the uncertainty around the PH assumption for OS (pg.114). The ERG's methodology uses a censor-at-switch analysis and does	It should be noted in the report that the ERG's preferred base case is unrealistic as it: - Extends OS but uses KM data for PFS and RFS leading to clinical implausibility. In addition, it can be seen that, even in isolation, the KM data for PFS and RFS is not	Given the ERG's stated uncertainty around the PH assumption with OS, the clinical implausibility of the ERG's modelling of OS in relation to PFS and RFS, removing subsequent therapy and the way the changes in the ERG's base case interact with each other, the ERG's methodology is misleading. Stating the ERG preferred base case	The ERG disagrees with Celgene's view. In response to each point in proposed amendments: ERG chose KM curves for PFS and RFS because they are a) as observed in the trial and b) non-parametric, i.e. free from risk of bias in fitting a parametric curve, which Celgene's own analysis identified as an issue. The resulting PFS and RFS

not allow for any subsequent therapy; this does not represent the view of the clinical expert consulted by the ERG, the data from the AZA-AML-001 trial or the clinician questionnaire used to calculate HRU by the Company. Pg. 118 of ERG report: text states that their OS modelling is "conservative" i.e. it is unrealistically increases the ICER, because it is biased in favour of AZA for OS. PFS and RFS are not increased (ERG uses KM data in their base case). The ERG's modelling of OS interacts with two other influential ERG changes 1. Scenario "E", the use of KM data only for RFS and PFS (which we have shown is inappropriate), and 2. Scenario "B", the calibration of relapse/PD HRU and costs. The unrealistic combination of these changes is misleading.	appropriate - Pushes patients into the PD disease state for longer which combines with other ERG changes such as equalised costs and HRU across all PD for both arms thus biasing against azacitidine - Suggests no subsequent therapy is received after 1st line therapy despite consistent evidence that this is unrealistic	without acknowledging its limitations and implausibility clearly is misleading to the reader.	data may have been affected by the use of subsequent therapies although it appears that at least for CCR arm subsequent azacitidine use occurred in few instances and their effect was negligible (this was confirmed by Celgene's response to ERG's questions for clarification). Acknowledging that the ERG's preferred base-case analysis results in a longer incremental period in relapse/PD under azacitidine than those of Celgene's base case, ERG have undertaken considerable sensitivity analyses. All of these show that the ICER is >£70,000 per QALY even when relapse/PD costs are zero. Please see our response to Issues 5 & 7. **No action taken.**

Issue 9 The use of KM curves from the trial for RFS and PFS within the model

<table>
<tr><th>Description of problem</th><th>Description of proposed amendment</th><th>Justification for amendment</th></tr>
<tr>
<td>
ERG preferred base case uses KM data for RFS and PFS. Section 5.4.2.4, page 128.

RFS and PFS KM curves, for all regimens, are not completely observed. The table below provides the final survival data from the observed KM curves; in the case of RFS in particular we would argue that some modelling is required.

Identifying the optimal fitting functional form is uncertain, however retaining KM data only underestimates benefit in all regimens.

The ERG's preferred modelling for OS increases OS compared with the Company's base case.

Survival rates in the KM curves are below:
<table>
<tr><th></th><th colspan="2">AZA</th><th colspan="2">CCR</th></tr>
<tr><th></th><th>Minimum PFS</th><th>Minimum RFS</th><th>Minimum PFS</th><th>Minimum RFS</th></tr>
<tr><td>All patients</td><td>0.0149</td><td>0.1390</td><td>0.0161</td><td>0.1014</td></tr>
<tr><td>IC</td><td>0.0938</td><td>0.1455</td><td>0.0588</td><td>0.1604</td></tr>
<tr><td>LDC</td><td>0.0362</td><td>0.1128</td><td>0.0242</td><td>0.0515</td></tr>
<tr><td>BSC</td><td>0.0909</td><td>0.2857</td><td>0.0588</td><td>N/a</td></tr>
<tr><td>Intermediate risk</td><td>0.0266</td><td>0.1581</td><td>0.0353</td><td>0.1039</td></tr>
<tr><td>Poor risk</td><td>0.0266</td><td>0.1581</td><td>0.0353</td><td>0.1039</td></tr>
<tr><td>With myelodysplasia-related changes</td><td>0.0845</td><td>0.2591</td><td>0.0484</td><td>0.0879</td></tr>
<tr><td>Without myelodysplasia-related changes</td><td>0.0155</td><td>0.0917</td><td>0.0245</td><td>0.1261</td></tr>
</table>
</td>
<td>
It should be noted in the report that KM data for RFS and PFS are not completely observed, and that some modelling is required.

It should be noted that the ERG's preferred base case increases OS, but does not increase PFS or RFS, and that there is no clinical reason for this.
</td>
<td>
It is factually inaccurate to suggest the KM data is complete for PFS and RFS.

It is misleading and clinically implausible to extend OS whilst not extending PFS and RFS.
</td>
</tr>
</table>

ERG response
The ERG accepts that its use of KM curves result in underestimation ('censoring') of time in the Remission state. However, this affects a minority of patients, i.e. those with a CR/CRi response, 28% in AZA and 25% in CCR, and given that the difference in censoring in AZA vs. CCR is 3.76 percentage points (0.1390 minus 0.1014) the bias against azacitidine is small and unlikely to be of significance to the results. **No action taken.**

Issue 10 The suggestion that EOL criteria are met based on restricted mean value

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
When assessing whether EOL criteria are met, on page 136, table 62 the ERG state "ERG analyses based on restricted mean survival at 30 months suggest an extension to life of 1.8–2.5 months" The ERG also state that "Extension to life should be assessed considering differences in mean overall survival in addition to median OS" The ERG also conclude (pg. 117) that the mean predicted difference in OS from their base case model is 3.64 months in favour of azacitidine.	It should be explicitly stated that using the restricted mean to determine the mean OS gain for the target population is inappropriate. The restricted mean underestimates the mean OS and is therefore not representative of the target population. It should also be stated that the median OS gain from the trial, as well as the mean OS gain from the cost effectiveness model indicate that EOL criteria are met.	It is factually inaccurate to state that the extension to OS is not met; the restricted mean is not suitable for basing the conclusion about the extension to life. The median OS gain from the AZA-AML-001 is 3.8 months and the modelled mean OS for corrections individually implemented by the ERG each produce a mean OS gain of >3 months. Additionally, the ERG conclude (pg. 117) that the mean predicted difference in OS from their base case model is 3.64 months in favour of azacitidine.	Overall survival curves have converged by 30 months in the ITT analysis (Celgene submission, Figure 8), and arguably also in the censor-at-switch analysis (Celgene submission, Figure 9). For this reason the difference in restricted mean survival (observed) is considered to be an unbiased estimate of the difference in unrestricted mean survival (unobserved); i.e., the restricted mean *is* a suitable basis for the conclusion about the extension to life. The ERG do not consider it necessary to state that restricted mean survival is an underestimate of unrestricted mean survival since this is widely understood and the ERG

			did not suggest anything to contradict this. The ERG note that even if median OS gain is preferred over mean OS gain (as estimated using restricted mean OS gain), the estimate is not robustly over 3 months, with a 95% CI of 1.0 to 6.5 months. **No action taken.**

Issue 11 Unexplained model results due to ERG changes

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
The combination of ERG changes do not explain changes in model results. Switching off all of the ERG changes results in an ICER of £36,931, not £20,648, the original base case. The ERG note on the worksheet "ERG changes" of the model that additional changes relate to Table 32 of the report; this table contains unit costs. It is not clear how the ERG have implemented their changes and why the ICER with ERG changes switched off is different to the Company's base case ICER.	More explanation of modifications made by the ERG	The ERG amendments cannot be checked for accuracy and the explanation does not seem to fit with the actual change in ICER. The ERG changes may interact with the changes that are listed. This can be seen by comparing Table 58 (piecewise changes) and Table 59 (cumulative changes). This interaction may be significant because of the impact in particular of the ERG's "A" and "B" amendments to the model.	Due to the high number of corrections made by the ERG to the model submitted by Celgene and due to the time constraints, option buttons were created only for major corrections. The list of corrections to the implementation of Celgene's model is shown in Table 57 of the ERG's report. An additional model correction, made by the ERG and discussed in section 5.3.11.2 of the report (p. 126), was implemented in the model as "MAX operator" option buttons.

			The duration of monitoring tests and transfusions implemented in the ERG's base case is described in section 6.1 (p. 134) of our report: "The base case analysis preferred by ERG included costs of monitoring tests and transfusions for the whole duration of the Remission and Non-remission phases of the model (as well as for the Relapse/PD phase)." In the executable model, all corrected cells are highlighted in yellow. **Action: ERG have produced a full listing of corrections to Celgene's model (below)**

Full listing of corrections made to Celgene's model

The ERG present below a full listing of corrections made to Celgene's model. Each correction is presented as an individual correction to Celgene's original model, i.e., there is no accumulation of corrections shown. Correct implementation of all of these corrections leads to analysis A (corrected base case). ICERs are shown for AZA versus CCR in the deterministic base case. These corrections are listed in order of decreasing ICER.

Correction 1

Issue	In the CCR arm, the patients receiving BSC only are assumed to incur drug administration costs in the Remission and Non-remission states, although the costs of administering BSC are not included after discontinuation of other active treatments until relapse/progression.	
Cells affected	'Model CCR'!AB23:AB153	
Original formula		**Revised formula**
Formula in 'Model CCR'!AB23: ='PSA inputs'!$D$71+K23*IF(C23<7,'PSA inputs'!$D$72,IF(C23>'PSA inputs'!$D$121,0,'PSA inputs'!$D$72))+MAX(L23-K23,0)*IF(C23<7,'PSA inputs'!$D$73,IF(C23>'PSA inputs'!$D$121,0,'PSA inputs'!$D$73))		replaced with: =('PSA inputs'!$D$71+K23*IF(C23<7,'PSA inputs'!$D$72,IF(C23>'PSA inputs'!$D$121,0,'PSA inputs'!$D$72))+MAX(L23-K23,0)*IF(C23<7,'PSA inputs'!$D$73,IF(C23>'PSA inputs'!$D$121,0,'PSA inputs'!$D$73))) * ('PSA inputs'!$D$16+'PSA inputs'!$D$17)
Formula in 'Model CCR'!AB24: =K24*IF(C24<7,'PSA inputs'!$D$72,IF(C24>'PSA inputs'!$D$121,0,'PSA inputs'!$D$72))+MAX(L24-K24,0)*IF(C24<7,'PSA inputs'!$D$73,IF(C24>'PSA inputs'!$D$121,0,'PSA inputs'!$D$73))		replaced with: =(K24*IF(C24<7,'PSA inputs'!$D$72,IF(C24>'PSA inputs'!$D$121,0,'PSA inputs'!$D$72))+MAX(L24-K24,0)*IF(C24<7,'PSA inputs'!$D$73,IF(C24>'PSA inputs'!$D$121,0,'PSA inputs'!$D$73))) * ('PSA inputs'!$D$16+'PSA inputs'!$D$17) (Correct the formulas in the cells below)
ICER (cost per QALY)	£43,676	

Correction 2

Issue	In the AZA and CCR arms, costs of tests and transfusions are not modelled for patients in the Relapse/progressive disease state.
Cells affected	'Model AZA'!AC23:AD153, 'Model CCR'!AC23:AD153
Original formula	**Revised formula**
Formula in 'Model AZA'!AC23: ='PSA inputs'!$D$57+K23*IF(C23<7,'PSA inputs'!$D$58,IF(C23>'PSA inputs'!$D$121,0,'PSA inputs'!$D$58))+MAX(L23-K23,0)*IF(C23<7,'PSA inputs'!$D$59,IF(C23>'PSA inputs'!$D$121,0,'PSA inputs'!$D$59))	replaced with: ='PSA inputs'!$D$57+K23*IF(C23<7,'PSA inputs'!$D$58,IF(C23>'PSA inputs'!$D$121,0,'PSA inputs'!$D$58))+MAX(L23-K23,0)*IF(C23<7,'PSA inputs'!$D$59,IF(C23>'PSA inputs'!$D$121,0,'PSA inputs'!$D$59)) +M23*'PSA inputs'!$D$60
Formula in 'Model AZA'!AC24: =K24*IF(C24<7,'PSA inputs'!$D$58,IF(C24>'PSA inputs'!$D$121,0,'PSA inputs'!$D$58))+MAX(L24-K24,0)*IF(C24<7,'PSA inputs'!$D$59,IF(C24>'PSA inputs'!$D$121,0,'PSA inputs'!$D$59))	replaced with: =K24*IF(C24<7,'PSA inputs'!$D$58,IF(C24>'PSA inputs'!$D$121,0,'PSA inputs'!$D$58))+MAX(L24-K24,0)*IF(C24<7,'PSA inputs'!$D$59,IF(C24>'PSA inputs'!$D$121,0,'PSA inputs'!$D$59)) +M24*'PSA inputs'!$D$60 (Correct the formulas in the cells below)
Formula in 'Model CCR'!AC23: ='PSA inputs'!$D$76+K23*IF(C23<7,'PSA inputs'!$D$77,IF(C23>'PSA inputs'!$D$121,0,'PSA inputs'!$D$77))+MAX(L23-K23,0)*IF(C23<7,'PSA inputs'!$D$78,IF(C23>'PSA inputs'!$D$121,0,'PSA inputs'!$D$78))	replaced with: ='PSA inputs'!$D$76+K23*IF(C23<7,'PSA inputs'!$D$77,IF(C23>'PSA inputs'!$D$121,0,'PSA inputs'!$D$77))+MAX(L23-K23,0)*IF(C23<7,'PSA inputs'!$D$78,IF(C23>'PSA inputs'!$D$121,0,'PSA inputs'!$D$78)) +M23*'PSA inputs'!$D$79
Formula in 'Model CCR'!AC24: =K24*IF(C24<7,'PSA inputs'!$D$77,IF(C24>'PSA inputs'!$D$121,0,'PSA inputs'!$D$77))+MAX(L24-K24,0)*IF(C24<7,'PSA inputs'!$D$78,IF(C24>'PSA inputs'!$D$121,0,'PSA inputs'!$D$78))	replaced with: =K24*IF(C24<7,'PSA inputs'!$D$77,IF(C24>'PSA inputs'!$D$121,0,'PSA inputs'!$D$77))+MAX(L24-K24,0)*IF(C24<7,'PSA inputs'!$D$78,IF(C24>'PSA inputs'!$D$121,0,'PSA inputs'!$D$78)) +M24*'PSA inputs'!$D$79 (Correct the formulas in the cells below)
Formula in 'Model AZA'!AD23: ='PSA inputs'!$D$62+K23*IF(C23<7,'PSA inputs'!$D$63,IF(C23>'PSA inputs'!$D$121,0,'PSA inputs'!$D$63))+MAX(L23-K23,0)*IF(C23<7,'PSA inputs'!$D$64,IF(C23>'PSA inputs'!$D$121,0,'PSA inputs'!$D$64))	replaced with: ='PSA inputs'!$D$62+K23*IF(C23<7,'PSA inputs'!$D$63,IF(C23>'PSA inputs'!$D$121,0,'PSA inputs'!$D$63))+MAX(L23-K23,0)*IF(C23<7,'PSA inputs'!$D$64,IF(C23>'PSA inputs'!$D$121,0,'PSA inputs'!$D$64)) +M23*'PSA inputs'!$D$65

Formula in 'Model AZA'!AD24: =K24*IF(C24<7,'PSA inputs'!$D$64,IF(C24>'PSA inputs'!$D$121,0,'PSA inputs'!$D$64))+MAX(L24-K24,0)*IF(C24<7,'PSA inputs'!$D$64,IF(C24>'PSA inputs'!$D$121,0,'PSA inputs'!$D$64))	replaced with: =K24*IF(C24<7,'PSA inputs'!$D$64,IF(C24>'PSA inputs'!$D$121,0,'PSA inputs'!$D$64))+MAX(L24-K24,0)*IF(C24<7,'PSA inputs'!$D$64,IF(C24>'PSA inputs'!$D$121,0,'PSA inputs'!$D$64)) +M24*'PSA inputs'!$D$65 (Correct the formulas in the cells below)
Formula in 'Model CCR'!AD23: ='PSA inputs'!$D$62+K23*IF(C23<7,'PSA inputs'!$D$82,IF(C23>'PSA inputs'!$D$121,0,'PSA inputs'!$D$82))+MAX(L23-K23,0)*IF(C23<7,'PSA inputs'!$D$83,IF(C23>'PSA inputs'!$D$121,0,'PSA inputs'!$D$83))	replaced with: ='PSA inputs'!$D$62+K23*IF(C23<7,'PSA inputs'!$D$82,IF(C23>'PSA inputs'!$D$121,0,'PSA inputs'!$D$82))+MAX(L23-K23,0)*IF(C23<7,'PSA inputs'!$D$83,IF(C23>'PSA inputs'!$D$121,0,'PSA inputs'!$D$83)) +M23*'PSA inputs'!$D$84
Formula in 'Model CCR'!AD24: =K24*IF(C24<7,'PSA inputs'!$D$83,IF(C24>'PSA inputs'!$D$121,0,'PSA inputs'!$D$83))+MAX(L24-K24,0)*IF(C24<7,'PSA inputs'!$D$83,IF(C24>'PSA inputs'!$D$121,0,'PSA inputs'!$D$83))	replaced with: =K24*IF(C24<7,'PSA inputs'!$D$83,IF(C24>'PSA inputs'!$D$121,0,'PSA inputs'!$D$83))+MAX(L24-K24,0)*IF(C24<7,'PSA inputs'!$D$83,IF(C24>'PSA inputs'!$D$121,0,'PSA inputs'!$D$83)) +M24*'PSA inputs'!$D$84 (Correct the formulas in the cells below)
ICER (cost per QALY)	£37,381

Correction 3

Issue	In the AZA and CCR arms, drug administration, monitoring tests and transfusion costs are double-counted during the 1st model cycle.
Cells affected	‘Model AZA ‘!AB23, AC23 and AD23; ‘Model CCR’! AB23, AC23 and AD23
Original formula	**Revised formula**
Formula in ‘Model AZA’!AB23: ='PSA inputs'!$D$52+K23*IF(C23<7,'PSA inputs'!$D$53,IF(C23>'PSA inputs'!$D$121,0,'PSA inputs'!$D$53))+MAX(L23-K23,0)*IF(C23<7,'PSA inputs'!$D$54,IF(C23>'PSA inputs'!$D$121,0,'PSA inputs'!$D$54))	replaced with: ='PSA inputs'!$D$52
Formula in ‘Model AZA’!AC23: ='PSA inputs'!$D$57+K23*IF(C23<7,'PSA inputs'!$D$58,IF(C23>'PSA inputs'!$D$121,0,'PSA inputs'!$D$58))+MAX(L23-K23,0)*IF(C23<7,'PSA inputs'!$D$59,IF(C23>'PSA inputs'!$D$121,0,'PSA inputs'!$D$59))	replaced with: ='PSA inputs'!$D$57
Formula in ‘Model AZA’!AD23: ='PSA inputs'!$D$62+K23*IF(C23<7,'PSA inputs'!$D$63,IF(C23>'PSA inputs'!$D$121,0,'PSA inputs'!$D$63))+MAX(L23-K23,0)*IF(C23<7,'PSA inputs'!$D$64,IF(C23>'PSA inputs'!$D$121,0,'PSA inputs'!$D$64))	replaced with: ='PSA inputs'!$D$62
Formula in ‘Model CCR’!AB23: ='PSA inputs'!$D$71+K23*IF(C23<7,'PSA inputs'!$D$72,IF(C23>'PSA inputs'!$D$121,0,'PSA inputs'!$D$72))+MAX(L23-K23,0)*IF(C23<7,'PSA inputs'!$D$73,IF(C23>'PSA inputs'!$D$121,0,'PSA inputs'!$D$73))	replaced with: ='PSA inputs'!$D$71
Formula in ‘Model CCR’!AC23: ='PSA inputs'!$D$76+K23*IF(C23<7,'PSA inputs'!$D$77,IF(C23>'PSA inputs'!$D$121,0,'PSA inputs'!$D$77))+MAX(L23-K23,0)*IF(C23<7,'PSA inputs'!$D$78,IF(C23>'PSA inputs'!$D$121,0,'PSA inputs'!$D$78))	replaced with: ='PSA inputs'!$D$76
Formula in ‘Model CCR’!AD23: ='PSA inputs'!$D$62+K23*IF(C23<7,'PSA inputs'!$D$82,IF(C23>'PSA inputs'!$D$121,0,'PSA inputs'!$D$82))+MAX(L23-K23,0)*IF(C23<7,'PSA inputs'!$D$83,IF(C23>'PSA inputs'!$D$121,0,'PSA inputs'!$D$83))	replaced with: ='PSA inputs'!$D$62
ICER (cost per QALY)	£35,532

Correction 4

Issue	In the CCR arm, transfusion costs in the Remission state from the 2nd cycle onwards are modelled using the transfusion costs for patients with stable disease.	
Cells affected	'Model CCR'!AD24:153	
Original formula		**Revised formula**
Formula: =K24*IF(C24<7,'PSA inputs'!$D$83,IF(C24>'PSA inputs'!$D$121,0,'PSA inputs'!$D$83))+MAX(L24-K24,0)*IF(C24<7,'PSA inputs'!$D$83,IF(C24>'PSA inputs'!$D$121,0,'PSA inputs'!$D$83))		replaced with: =K24*IF(C24<7,'PSA inputs'!$D$82,IF(C24>'PSA inputs'!$D$121,0,'PSA inputs'!$D$82))+MAX(L24-K24,0)*IF(C24<7,'PSA inputs'!$D$83,IF(C24>'PSA inputs'!$D$121,0,'PSA inputs'!$D$83))
ICER (cost per QALY)	£28,147	

Correction 5

Issue	In the AZA and CCR arms, monitoring tests and transfusions were modelled only while the patient remained on their initial treatment.
Cells affected	'Model AZA'!AC24:AD153; 'Model CCR'!AC24:AD153
Original formula	**Revised formula**
Formula in 'Model AZA'!AC24: =K24*IF(C24<7,'PSA inputs'!$D$58,IF(C24>'PSA inputs'!$D$121,0,'PSA inputs'!$D$58))+MAX(L24-K24,0)*IF(C24<7,'PSA inputs'!$D$59,IF(C24>'PSA inputs'!$D$121,0,'PSA inputs'!$D$59))	replaced with: =K24*'PSA inputs'!$D$58+MAX(L24-K24,0)*'PSA inputs'!$D$59 (Correct the formulas in the cells below)
Formula in 'Model AZA'!AD24: =K24*IF(C24<7,'PSA inputs'!$D$64,IF(C24>'PSA inputs'!$D$121,0,'PSA inputs'!$D$64))+MAX(L24-K24,0)*IF(C24<7,'PSA inputs'!$D$64,IF(C24>'PSA inputs'!$D$121,0,'PSA inputs'!$D$64))	replaced with: =K24*'PSA inputs'!$D$64+MAX(L24-K24,0)*'PSA inputs'!$D$64 (Correct the formulas in the cells below)
Formula in 'Model CCR'!AC24: =K24*IF(C24<7,'PSA inputs'!$D$77,IF(C24>'PSA inputs'!$D$121,0,'PSA inputs'!$D$77))+MAX(L24-K24,0)*IF(C24<7,'PSA inputs'!$D$78,IF(C24>'PSA inputs'!$D$121,0,'PSA inputs'!$D$78))	replaced with: =K24*'PSA inputs'!$D$77+MAX(L24-K24,0)*'PSA inputs'!$D$78 (Correct the formulas in the cells below)
Formula in 'Model CCR'!AD24: =K24*IF(C24<7,'PSA inputs'!$D$83,IF(C24>'PSA inputs'!$D$121,0,'PSA inputs'!$D$83))+MAX(L24-K24,0)*IF(C24<7,'PSA inputs'!$D$83,IF(C24>'PSA inputs'!$D$121,0,'PSA inputs'!$D$83))	replaced with: =K24*'PSA inputs'!$D$83+MAX(L24-K24,0)*'PSA inputs'!$D$83 The wrong cell reference in this formula is discussed above. (Correct the formulas in the cells below)
ICER (cost per QALY)	£26,696

Correction 6

Issue	In the CCR arm, the patients receiving LDAC have costs of drug administration, tests and transfusions estimated based on the treatment duration of azacitidine, rather than the treatment duration of LDAC.	
Cells affected	'Model CCR'!AB23:AD153	
Original formula		**Revised formula**
Cell reference: 'PSA inputs'!$D$121		replaced with: 'PSA inputs'!$D$158
ICER (cost per QALY)	£26,537	

Correction 7

Issue	In the CCR arm, the patients receiving IC are assumed to incur drug administration costs after their treatment is discontinued (i.e., after cycle 2).	
Cells affected	'Model CCR'!AB25:AB153	
Original formula		**Revised formula**
Formula in 'Model CCR'!AB25: =K25*IF(C25<7,'PSA inputs'!$D$72,IF(C25>'PSA inputs'!$D$121,0,'PSA inputs'!$D$72))+MAX(L25-K25,0)*IF(C25<7,'PSA inputs'!$D$73,IF(C25>'PSA inputs'!$D$121,0,'PSA inputs'!$D$73))		replaced with: =(K25*IF(C25<7,'PSA inputs'!$D$72,IF(C25>'PSA inputs'!$D$121,0,'PSA inputs'!$D$72))+MAX(L25-K25,0)*IF(C25<7,'PSA inputs'!$D$73,IF(C25>'PSA inputs'!$D$121,0,'PSA inputs'!$D$73)))* ('PSA inputs'!$D$17+'PSA inputs'!$D$18) (Correct the formulas in the cells below)
ICER (cost per QALY)	£26,333	

Correction 8

Issue	In the CCR arm, transfusion costs in the pre-response period were modelled using transfusion costs for patients receiving azacitidine.	
Cells affected	'Model CCR'!AD23	
Original formula		**Revised formula**
Cell reference: 'PSA inputs'!$D$62		replaced with: 'PSA inputs'!$D$81
ICER (cost per QALY)	£21,742	

Correction 9

Issue	In the CCR arm, the daily dose of cytarabine for patients receiving LDAC is estimated assuming calculations using units of mg/m²/day, whereas the parameter used in AZA-AML-001 has units of mg/day.	
Cells affected	'HRU costs'!C115:117	
Original formula		**Revised formula**
Formula: ='PSA inputs'!D15*'PSA inputs'!D156		replaced with: 'PSA inputs'!D156
ICER (cost per QALY)	£20,648 (only affects scenario analyses)	

Correction 10

Issue	In the AZA and CCR arms, the formulae for calculating wastage with 30% tolerance (used in a scenario analysis) are incorrect.
Cells affected	'HRU costs'!E103, E109, E113, E117, E123, E127, E133, E137
Original formula	**Revised formula**
Formula in 'HRU costs'!E103: =IF(E102*(1-'Drug costs'!$L$14)<E101,E101,(E102*(1-'Drug costs'!$L$14)))	replaced with: = 'Drug costs'!$L$14*E101+(1-'Drug costs'!$L$14)*E102
Formula in 'HRU costs'!E109: =IF(E108*(1-'Drug costs'!$L$14)<E107,E107,(E108*(1-'Drug costs'!$L$14)))	replaced with: ='Drug costs'!$L$14*E107+(1-'Drug costs'!$L$14)*E108
Formula in 'HRU costs'!E113: =IF(E112*(1-'Drug costs'!$L$14)<E111,E111,(E112*(1-'Drug costs'!$L$14)))	replaced with: ='Drug costs'!$L$14*E111+(1-'Drug costs'!$L$14)*E112
Formula in 'HRU costs'!E117: =IF(E116*(1-'Drug costs'!$L$14)<E115,E115,(E116*(1-'Drug costs'!$L$14)))	replaced with: ='Drug costs'!$L$14*E115+(1-'Drug costs'!$L$14)*E116
Formula in 'HRU costs'!E123: =IF(E122*(1-'Drug costs'!$L$14)<E121,E121,(E122*(1-'Drug costs'!$L$14)))	replaced with: = 'Drug costs'!$L$14*E121+(1-'Drug costs'!$L$14)*E122
Formula in 'HRU costs'!E127: =IF(E126*(1-'Drug costs'!$L$14)<E125,E125,(E126*(1-'Drug costs'!$L$14)))	replaced with: ='Drug costs'!$L$14*E125+(1-'Drug costs'!$L$14)*E126
Formula in 'HRU costs'!E133: =IF(E132*(1-'Drug costs'!$L$14)<E131,E131,(E132*(1-'Drug costs'!$L$14)))	replaced with: ='Drug costs'!$L$14*'HRU costs'!E131+(1-'Drug costs'!$L$14)*'HRU costs'!E132
Formula in 'HRU costs'!E137: =IF('ERG changes'!V10 = 2,('Drug costs'!$L$14*E135+(1-'Drug costs'!$L$14)*E136),IF(E136*(1-'Drug costs'!$L$14)<E135,E135,(E136*(1-'Drug costs'!$L$14))))	replaced with: 'Drug costs'!$L$14*E135+(1-'Drug costs'!$L$14)*E136
ICER (cost per QALY)	£20,648 (only affects scenario analyses)

Correction 11

Issue	In the AZA and CCR arms, the Kaplan-Meier curves for OS, PFS and RFS (for use in scenario analyses) are incorrectly referenced to the curves for patients with IC as their pre-specified CCR.
Cells affected	'KM data'!DD6:DI49
Original formula	**Revised formula**
Formula in 'KM data'!DD6: =IF((CHOOSE(PatientGroup,BF6,BL6,BR6,BX6,CD6,CJ6,CP6))=0, NA, CHOOSE(PatientGroup,BF6,BL6,BR6,BX6,CD6,CJ6,CP6))	replaced with: =IF((CHOOSE(PatientGroup, AZ6,BF6,BL6,BR6,BX6,CD6,CJ6,CP6))=0, NA, CHOOSE(PatientGroup, AZ6,BF6,BL6,BR6,BX6,CD6,CJ6,CP6)) (Correct the formulas in the cells below)
Formula in 'KM data'!DE6: =IF((CHOOSE(PatientGroup,BG6,BM6,BS6,BY6,CE6,CK6,CQ6))=0, NA, CHOOSE(PatientGroup,BG6,BM6,BS6,BY6,CE6,CK6,CQ6))	replaced with: =IF((CHOOSE(PatientGroup, BA6,BG6,BM6,BS6,BY6,CE6,CK6,CQ6))=0, NA, CHOOSE(PatientGroup, BA6,BG6,BM6,BS6,BY6,CE6,CK6,CQ6)) (Correct the formulas in the cells below)
Formula in 'KM data'!DF6: =IF((CHOOSE(PatientGroup,BH6,BN6,BT6,BZ6,CF6,CL6,CR6))=0, NA, CHOOSE(PatientGroup,BH6,BN6,BT6,BZ6,CF6,CL6,CR6))	replaced with: =IF((CHOOSE(PatientGroup,BB6,BH6,BN6,BT6,BZ6,CF6,CL6,CR6))=0, NA, CHOOSE(PatientGroup,BB6,BH6,BN6,BT6,BZ6,CF6,CL6,CR6)) (Correct the formulas in the cells below)
Formula in 'KM data'!DG6: =IF((CHOOSE(PatientGroup,BI6,BO6,BU6,CA6,CG6,CM6,CS6))=0, NA, CHOOSE(PatientGroup,BI6,BO6,BU6,CA6,CG6,CM6,CS6))	replaced with: =IF((CHOOSE(PatientGroup, BC6,BI6,BO6,BU6,CA6,CG6,CM6,CS6))=0, NA, CHOOSE(PatientGroup, BC6,BI6,BO6,BU6,CA6,CG6,CM6,CS6)) (Correct the formulas in the cells below)
Formula in 'KM data'!DH6: =IF((CHOOSE(PatientGroup,BJ6,BP6,BV6,CB6,CH6,CN6,CT6))=0, NA, CHOOSE(PatientGroup,BJ6,BP6,BV6,CB6,CH6,CN6,CT6))	replaced with: =IF((CHOOSE(PatientGroup, BD6,BJ6,BP6,BV6,CB6,CH6,CN6,CT6))=0, NA, CHOOSE(PatientGroup, BD6,BJ6,BP6,BV6,CB6,CH6,CN6,CT6)) (Correct the formulas in the cells below)

Formula in 'KM data'!DI6: =IF((CHOOSE(PatientGroup,BK6,BQ6,BW6,CC6,CI6,CO6,CU6))=0, NA, CHOOSE(PatientGroup,BK6,BQ6,BW6,CC6,CI6,CO6,CU6))		replaced with: =IF((CHOOSE(PatientGroup, BE6,BK6,BQ6,BW6,CC6,CI6,CO6,CU6))=0, NA, CHOOSE(PatientGroup, BE6,BK6,BQ6,BW6,CC6,CI6,CO6,CU6)) (Correct the formulas in the cells below)
ICER (cost per QALY)	£20,648 (only affects scenario analyses)	

Correction 12

Issue	In the AZA and CCR arms, the costs of drug administration, monitoring tests and transfusions for patients in the Non-remission (stable disease) state were calculated assuming the proportion of such patients in the cohort is estimated by $PFS \times p_{SD} - RFS \times p_{Response}$ (or zero if this is negative) where $PFS \times p_{SD}$ is the correct formula, $p_{Response}$ is the proportion of patients with CR/CRi response and $pSD = 1 - p_{Response}$.
Cells affected	"Model AZA"! AB23:AD153; "Model CCR"! AB23:AD153

Original formula	**Revised formula**
Formula in 'Model AZA'!AB23: ='PSA inputs'!$D$52+K23*IF(C23<7,'PSA inputs'!$D$53,IF(C23>'PSA inputs'!$D$121,0,'PSA inputs'!$D$53))+MAX(L23-K23,0)*IF(C23<7,'PSA inputs'!$D$54,IF(C23>'PSA inputs'!$D$121,0,'PSA inputs'!$D$54))	replaced with: ='PSA inputs'!$D$52+K23*IF(C23<7,'PSA inputs'!$D$53,IF(C23>'PSA inputs'!$D$121,0,'PSA inputs'!$D$53))+L23*IF(C23<7,'PSA inputs'!$D$54,IF(C23>'PSA inputs'!$D$121,0,'PSA inputs'!$D$54))
Formula in 'Model AZA'!AB24: =K24*IF(C24<7,'PSA inputs'!$D$53,IF(C24>'PSA inputs'!$D$121,0,'PSA inputs'!$D$53))+MAX(L24-K24,0)*IF(C24<7,'PSA inputs'!$D$54,IF(C24>'PSA inputs'!$D$121,0,'PSA inputs'!$D$54))	replaced with: =K24*IF(C24<7,'PSA inputs'!$D$53,IF(C24>'PSA inputs'!$D$121,0,'PSA inputs'!$D$53))+L24*IF(C24<7,'PSA inputs'!$D$54,IF(C24>'PSA inputs'!$D$121,0,'PSA inputs'!$D$54)) (Correct the formulas in the cells below)
Formula in 'Model CCR'!AB23: ='PSA inputs'!$D$71+K23*IF(C23<7,'PSA inputs'!$D$72,IF(C23>'PSA inputs'!$D$121,0,'PSA inputs'!$D$72))+MAX(L23-K23,0)*IF(C23<7,'PSA inputs'!$D$73,IF(C23>'PSA inputs'!$D$121,0,'PSA inputs'!$D$73))	replaced with: ='PSA inputs'!$D$71+K23*IF(C23<7,'PSA inputs'!$D$72,IF(C23>'PSA inputs'!$D$121,0,'PSA inputs'!$D$72))+L23*IF(C23<7,'PSA inputs'!$D$73,IF(C23>'PSA inputs'!$D$121,0,'PSA inputs'!$D$73))
Formula in 'Model CCR'!AB24: =K24*IF(C24<7,'PSA inputs'!$D$72,IF(C24>'PSA inputs'!$D$121,0,'PSA inputs'!$D$72))+MAX(L24-K24,0)*IF(C24<7,'PSA inputs'!$D$73,IF(C24>'PSA inputs'!$D$121,0,'PSA inputs'!$D$73))	replaced with: =K24*IF(C24<7,'PSA inputs'!$D$72,IF(C24>'PSA inputs'!$D$121,0,'PSA inputs'!$D$72))+L24*IF(C24<7,'PSA inputs'!$D$73,IF(C24>'PSA inputs'!$D$121,0,'PSA inputs'!$D$73)) (Correct the formulas in the cells below)
Formula in 'Model AZA'!AC23: ='PSA inputs'!$D$57+K23*IF(C23<7,'PSA inputs'!$D$58,IF(C23>'PSA inputs'!$D$121,0,'PSA inputs'!$D$58))+MAX(L23-K23,0)*IF(C23<7,'PSA inputs'!$D$59,IF(C23>'PSA inputs'!$D$121,0,'PSA inputs'!$D$59))	replaced with: ='PSA inputs'!$D$57+K23*IF(C23<7,'PSA inputs'!$D$58,IF(C23>'PSA inputs'!$D$121,0,'PSA inputs'!$D$58))+L23*IF(C23<7,'PSA inputs'!$D$59,IF(C23>'PSA inputs'!$D$121,0,'PSA inputs'!$D$59))

Formula in ‘Model AZA’!AC24: =K24*IF(C24<7,'PSA inputs'!$D$58,IF(C24>'PSA inputs'!$D$121,0,'PSA inputs'!$D$58))+MAX(L24-K24,0)*IF(C24<7,'PSA inputs'!$D$59,IF(C24>'PSA inputs'!$D$121,0,'PSA inputs'!$D$59))	replaced with: =K24*IF(C24<7,'PSA inputs'!$D$58,IF(C24>'PSA inputs'!$D$121,0,'PSA inputs'!$D$58))+L24*IF(C24<7,'PSA inputs'!$D$59,IF(C24>'PSA inputs'!$D$121,0,'PSA inputs'!$D$59)) (Correct the formulas in the cells below)
Formula in ‘Model CCR’!AC23: ='PSA inputs'!$D$76+K23*IF(C23<7,'PSA inputs'!$D$77,IF(C23>'PSA inputs'!$D$121,0,'PSA inputs'!$D$77))+MAX(L23-K23,0)*IF(C23<7,'PSA inputs'!$D$78,IF(C23>'PSA inputs'!$D$121,0,'PSA inputs'!$D$78))	replaced with: ='PSA inputs'!$D$76+K23*IF(C23<7,'PSA inputs'!$D$77,IF(C23>'PSA inputs'!$D$121,0,'PSA inputs'!$D$77))+L23*IF(C23<7,'PSA inputs'!$D$78,IF(C23>'PSA inputs'!$D$121,0,'PSA inputs'!$D$78))
Formula in ‘Model CCR’!AC24: =K24*IF(C24<7,'PSA inputs'!$D$77,IF(C24>'PSA inputs'!$D$121,0,'PSA inputs'!$D$77))+MAX(L24-K24,0)*IF(C24<7,'PSA inputs'!$D$78,IF(C24>'PSA inputs'!$D$121,0,'PSA inputs'!$D$78))	replaced with: =K24*IF(C24<7,'PSA inputs'!$D$77,IF(C24>'PSA inputs'!$D$121,0,'PSA inputs'!$D$77))+L24*IF(C24<7,'PSA inputs'!$D$78,IF(C24>'PSA inputs'!$D$121,0,'PSA inputs'!$D$78)) (Correct the formulas in the cells below)
Formula in ‘Model AZA’!AD23: ='PSA inputs'!$D$62+K23*IF(C23<7,'PSA inputs'!$D$63,IF(C23>'PSA inputs'!$D$121,0,'PSA inputs'!$D$63))+MAX(L23-K23,0)*IF(C23<7,'PSA inputs'!$D$64,IF(C23>'PSA inputs'!$D$121,0,'PSA inputs'!$D$64))	replaced with: ='PSA inputs'!$D$62+K23*IF(C23<7,'PSA inputs'!$D$63,IF(C23>'PSA inputs'!$D$121,0,'PSA inputs'!$D$63))+L23*IF(C23<7,'PSA inputs'!$D$64,IF(C23>'PSA inputs'!$D$121,0,'PSA inputs'!$D$64))
Formula in ‘Model AZA’!AD24: =K24*IF(C24<7,'PSA inputs'!$D$64,IF(C24>'PSA inputs'!$D$121,0,'PSA inputs'!$D$64))+MAX(L24-K24,0)*IF(C24<7,'PSA inputs'!$D$64,IF(C24>'PSA inputs'!$D$121,0,'PSA inputs'!$D$64))	replaced with: =K24*IF(C24<7,'PSA inputs'!$D$64,IF(C24>'PSA inputs'!$D$121,0,'PSA inputs'!$D$64))+L24*IF(C24<7,'PSA inputs'!$D$64,IF(C24>'PSA inputs'!$D$121,0,'PSA inputs'!$D$64)) (Correct the formulas in the cells below)
ICER (cost per QALY)	−£25,485 (AZA dominant)