ADDENDUM TO

Azacitidine for the treatment of myelodysplastic syndrome, chronic myelomonocytic leukaemia and acute myeloid leukaemia

Produced by West Midlands Health Technology Assessment Collaboration
Unit of Public Health, Epidemiology and Biostatistics
University of Birmingham
Edgbaston
Birmingham B15 2TT UK

Authors
Richard Edlin, Lecturer in Health Economics[2]
Martin Connock, Systematic Reviewer[1]
Sandy Tubeuf, Research Fellow[2]
Jeff Round, Lecturer[2]
Anne Fry-Smith, Information Specialist[1]
Chris Hyde, Senior Lecturer[1]
Wendy Greenheld, Systematic Reviewer[1]

[1]Unit of Public Health, Epidemiology and Biostatistics, University of Birmingham
[2]Academic Unit of Health Economics, Institute of Health Sciences, University of Leeds
Centre for Clinical Haematology, Queen Elizabeth Hospital, University Hospitals Birmingham NHS Foundation Trust

Correspondence to Chris Hyde, Unit of Public Health, Epidemiology and Biostatistics, University of Birmingham

Date completed June 2009

Source of funding: This report was commissioned by the NIHR HTA Programme as project number 08/84/01.

Declared competing interests of the authors: No competing interests – Hyde, Connock, Round, Tubeuf, Edlin, Fry-Smith, Greenheld, Craddock

Acknowledgements Professor C Craddock (Centre for Clinical Haematology, Queen Elizabeth Hospital, University Hospitals Birmingham NHS Foundation Trust) provided clinical advice and comments on the clinical aspects of the report.

1 Scope of the ADDENDUM

This addendum addresses and, as far as possible within time constraints attempts to resolve, issues relating to

- deficiencies identified in the submitted economic model.
- concerns associated with survival inputs into the economic model.

2 Concerns regarding survival inputs to economic model

The original ERG report commented: "*The credibility of the estimates of cost-effectiveness is the pre-dominant issue*", and "*even if the model is corrected further areas of uncertainty will remain*".

Regarding such areas of remaining uncertainty there were three main concerns

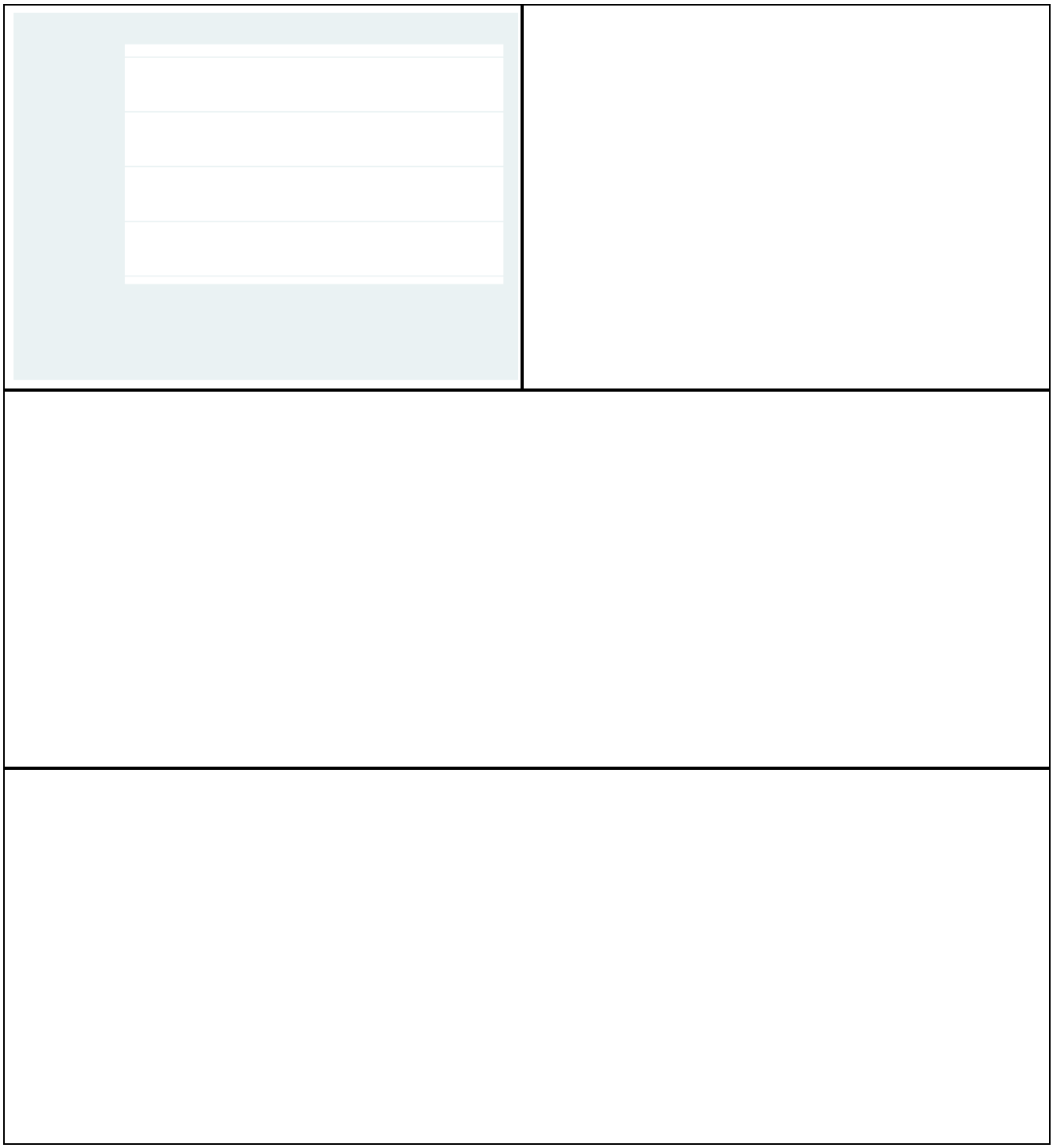
- Lack in the MS of quality information about overall survival of patients in the subgroups analysed in the economic model.
- A lack of face validity in the modelling of extrapolated overall survival.
- Inability to validate the survival in the AML state employed in the economic analyses.

These are addressed in this part of the addendum: Section 2.1 considers the MS provision of survival data for the subgroups that were input to the economic model. Section 2.2 considers the face validity of the log-logistic fits for overall survival that were used in the base case for the economic model and discusses the merit of potential alternative fits. Section 2.3 considers the data for survival in the AML state and describes the difficulties encountered in attempting to validate the model input using the observed data nested in the economic model.

2.1 Overall survival in the subgroups analysed in the economic model

The economic analysis considered three subgroups separately (BSC, LDC and SDC). Although the MS provided graphs showing survival for subgroups (Fig 7.2 a-f) these lacked information on the number of patients at risk and provided no indication

of the uncertainty (95% CI) inherent in the observed data; the latter deficiency also applied to MS presentation of survival for azacitidine and CCR groups as a whole (MS Fig 6.5). The figure below provides these details (data was extracted from the manufacturer's model; further details are provided in Appendix 1.a).



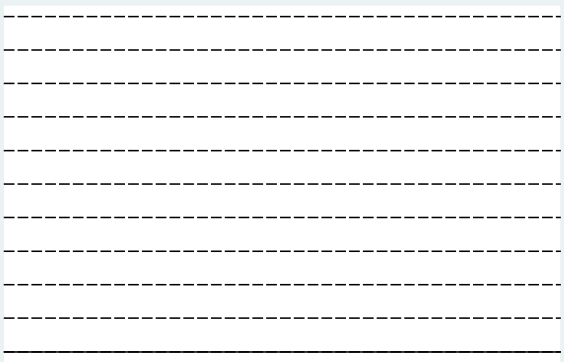
Of note is the considerable uncertainty associated with survival in the SDC subgroups; statistically significant superiority for azacitidine versus SDC was not established (95% CI for HR 0.33 to 1.74). The economic model generated the most favourable ICER for this subgroup yet convincing evidence of effectiveness was lacking.

2.2 Face validity of log-logistic fits for modelled overall survival

The manufacturer's base case economic analysis employed log-logistic fits to observed survival in the six subgroups (BSC-aza, BSC, LDC-aza, LDC, SDC-aza and SDC) and extrapolated the fits to a 25 year time horizon. In sensitivity analysis the alternative of Weibull fits for overall survival were explored.

Although MS (Fig 7.2 a to f) shows log-logistic and Weibull fits to the empirical data the extrapolation in the figure is only extended to 80 cycles (7.7 years) and little sense of the flattening out of these curves is conveyed. To remedy this deficiency ERG have extended the extrapolations to 30 years and superimposed these on Kaplan-Meier plots of the empirical data. The resulting graphs for each subgroup are shown in the figure below. In each case the lower curve represents the Weibull fit.

When the model log-logistic fits to overall survival data are extrapolated to 30 years about 4 or 5% of patients in the azacitidine subgroups are still alive. Taking the mean age reported for each subgroup and adding 30 years indicates that an appreciable number of azacitidine MDS patients would survive into their tenth decade. This considerable azacitidine-dependent improvement to the generally accepted prognosis for MDS patients was not observed when Weibull distributions are fitted. Clinical opinion may consider that the Weibull extrapolation has greater face validity and may be more suitable for the base case economic analysis.



Because of the considerable flattening of the log-logistic fitted survival curves seen at long extrapolation times the ERG asked the question:

- what is the modelled prediction for survival for 1% of the BSC patients treated with azacitidine?

$$S(t) = 1 / (1+((p*t)^{\lambda}))$$

Solving the equation above for t when S(t) = 0.01 using the values for the parameters P and λ defined in the model (0.048 and 1.152 respectively) yielded a value of 1125 cycles which is approximately 108 years.

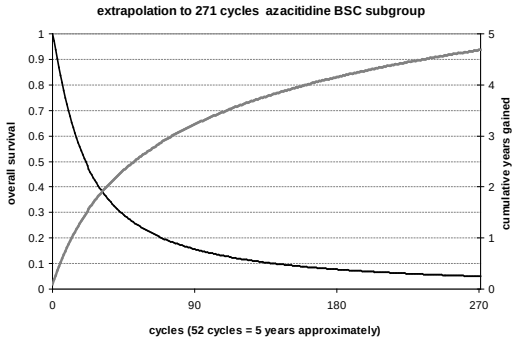
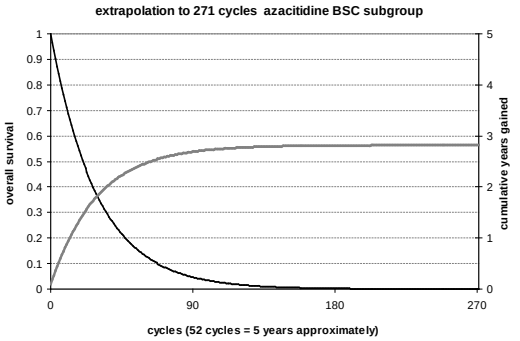
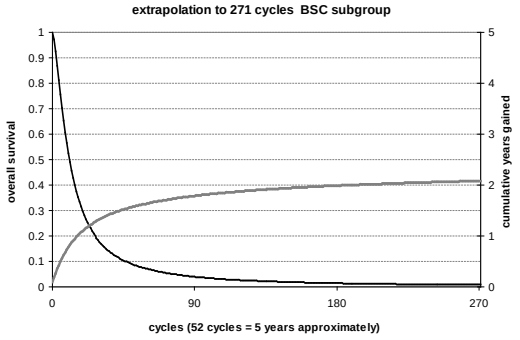
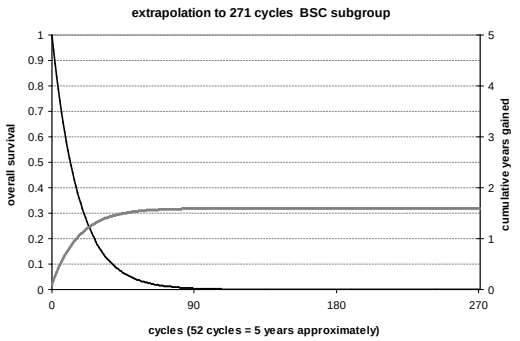
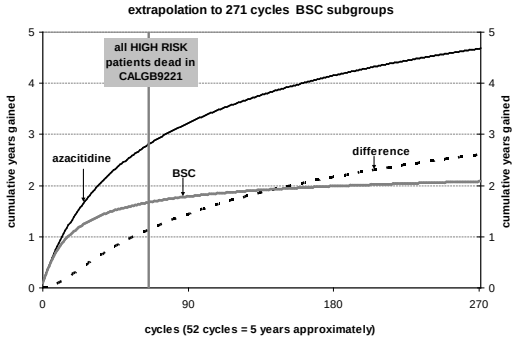
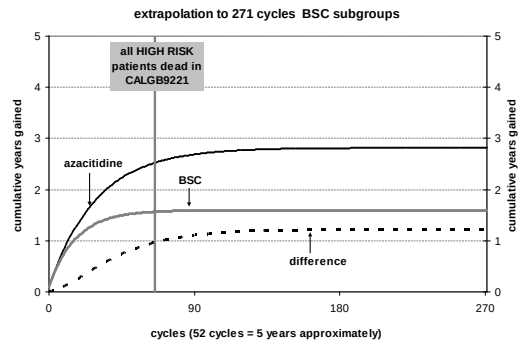
Because the MS reported only on log-logistic and Weibull fits to the empirical overall survival data ERG explored alternative fits.

ERG extracted data from the manufacturer's model and fitted exponential, Weibull, Gompertz, lognormal and log-logistic distributions using STATA software. AIC values were calculated for each fit (table below, for each subgroup the lowest AIC value is shaded). Although differences in goodness of fit were not great, according to AIC criteria an exponential fit was superior to other fits in four of the six subgroups and a log-logistic fit was superior in none. The choice of log-logistic fits for base case analysis is not well supported and the choice of exponential fits is defendable.

	Overall survival (Pre-selected Subgroups)					
Fitted Distribution	Azacitidine-BSC AIC value	BSC AIC value	Azacitidine-LDC AIC value	LDC AIC value	Azacitidine-SDC AIC value	SDC AIC value
Exponential	282.0821	276.58	111.859432	130.7675	47.67193	55.11062
Weibull	284.0715	277.302	113.75213	131.8855	49.66875	51.44694
Gompertz	283.8299	278.464	113.733154	132.1136	49.66913	53.44262
Lognormal	284.1122	270.196	112.81171	131.3413	50.21971	50.24947
Log-logistic	283.7325	271.382	113.605622	132.7571	49.97238	51.19378

A problem with the log-logistic fits when extrapolated over the 271 cycle time-horizon of the economic model is that azacitidine apparently generates considerable survival benefit well after patients with this severity of disease would be expected to have died (see for example the survival observed for high risk patients in the trial CALBG 9221 shown in Fig A2.1 in the manufacturer's 8 April 2008 response to request for clarification). In contrast exponential fits produce survival benefit from azacitidine that

is more reasonable in that most benefit is generated within the expected life span of patients and is consistent with the proposed mode of action of azacitidine as an agent that delays progression and death by an median of about nine months (difference in median survival times, MS table 6.7 page 42). These differences are illustrated for the BSC subgroups in the graphs below (left log-logistic fits, right exponential fits).

Log-logistic	Exponential
	
	
	

It is noticeable that the log-logistic gain in survival benefit for the azacitidine group compared to the BSC subgroup is still increasing after 270 cycles. Because of the severe flattening of the log-logistic fit for the azacitidine group at extended times a

survival gain for this group continues to accumulate well beyond 3000 cycles (about 287 years).

2.3 Survival in the AML state

Time to progression (TTP) to AML was modelled by back calculation using a pooled estimate of survival in AML (median 3.65 cycles, mean of 6.67 cycles). The manufacturer's responses to requests for clarification about survival in AML did not provide further information. ERG therefore undertook analyses to validate the median survival in AML, test the assumption that patient subgroups experienced the same survival in AML, compare modelled with observed survival in AML, and examine the reliability of the back calculation to TTP.

The results from these analyses are summarised below and indicate that the estimates of survival in AML are potentially unreliable and that the results of the modelling procedure should be viewed with considerable caution.

2.3.1 Replication of median survival in the AML state

One issue associated with the stated pooled estimate for survival in AML was the inability to replicate the median value of 3.65 cycles. The ERG extracted data for survival in AML from the model and generated the Kaplan-Meier plot shown below left. The median survival was 5.69 time units. A difficulty arose from the model's inconsistent labelling of time in months or cycles, so that the median could represent 5.69 cycles, or if time units are months, 4.94 cycles.[1]

[1] Note: model data listed two patients with event at t=0, these are excluded in analysis leaving 133 at risk not 135

The discrepancy from 3.65 cycles may be due to censoring. If no censoring is implemented the K-M plot above right is generated. The median time for this plot is 3.91 cycles or, with time units as months, 3.4 cycles. Without censoring the median represents the median time experienced in AML during the trial. An alternative procedure is to include all 358 patients in the analysis and censor all patients that did not progress; in this analysis a median time was not reached.

2.3.2 Observed time in AML state

The MS (page 70) stated that clinicians considered the time spent in AML would not differ between the treatment arms of the AZA-001 trial. How the clinical advice was obtained and how many clinicians were involved was unclear. No empirical data was presented in the MS that examined this assumption for the investigator-selected subgroups. Therefore data was extracted and K-M plots constructed (see appendix 1.b).

Hazard ratios (aza versus CCR) did not reach statistical significance. There were differences in observed median survival between aza and CCR within and between subgroups. There was little difference in median survival between all AZ and all CCR patients. The results are summarised below.

Patients / subgroups	Median (cycles) azacitidine	Median (cycles) CCR	HR aza v. CCR	95% CI HR	p
All			0.932	0.615 to 1.413	0.740
BSC	7.29	6.14	0.777	0.4500 to 1.341	0.910
LDC	2.97	8.09	1.424	0.6051 to 3.351	0.418
SDC	1.71	5.69	1.939	0.6411 to 5.862	0.241

Potential bias in estimating the start of the AML state (see below) means these results should be viewed with caution.

The modelled TTP adopted a log-logistic distribution for survival in AML. However this was not well supported from observed survival in the investigator-selected subgroups (see appendix 1.c).

2.3.3 Observed versus modelled time in AML

A comparison was made between observed and modelled survival in AML.

The manufacturer's modelled time in AML (median 3.65 cycles) used the log-logistic shape parameter (λ) for overall survival (page 81) and a scale parameter of p = 1/3.65 (p = 1/median). This generates the survival curves shown below for aza and CCR patients that progressed. After relatively few cycles modelled and observed survival curves depart substantially. Both observed and modelled survival curves are at risk of potential bias (see below).

An alternative procedure briefly described in the manufacturer's response (29 May 2009) to the second request for clarification apparently offsets the curve by 7 cycles from the overall survival curve.

Elsewhere in the MS (page 78) a pooled AML mortality rate of 0.135/cycle is described. This rate defines an exponential survival relationship with a median of 5.1 cycles (median = [ln 0.5] / [-0.135]). This is in contrast to, and inconsistent with, the 3.65 cycle median (log-logistic distribution) that was used to model TTP.

2.3.4 Use of time in AML rather than observed time to progression

Rather than use observed time to progression to AML (TTP) the model used pooled survival in AML to back calculate TTP. One reason given for this was:

- There were so many censorings in the observed TTP analysis that no reliable estimate could be obtained for subgroups.

If excess censoring resulted in unreliable estimates of TTP it must be assumed that this also holds for time spent in AML since a good estimate of the median time in

AML depends on reliable identification of patients who progressed and of the time of their progression. Thus the modelling procedure adopted merely mirrors and does not remedy the deficiency of a reliable estimate of TTP.

2.3.5 The reason observed TTP was not used in the model

The MS and the AZA-001 trial publication reported a longer median TTP for aza-treated patients than CCR patients (median 17.8 and 11.5 months respectively; HR 0.5 95%CI 0.35 to 0.70). In this analysis patients were censored at the time of their last bone marrow assessment rather than at end of follow up. Consequently, if bone marrow assessments were [redacted] throughout the trial duration, or were [redacted] relative to the other, there is a potential risk of bias in the estimate of TTP and in comparison between treatments. The CIC AZA-001 trial report contained information relevant to this issue and indicated that [redacted]% and [redacted]% of expected assessments were [redacted] respectively, and that the [redacted] was likely [redacted] (information summarised in appendix 1d).

In summary there is uncertainty and risk of bias associated with the estimate of TTP and comparisons between treatments that is likely to equally affect an estimate of survival in the AML state.

3 Economic evaluation

The submitted economic evaluation is a cost–utility analysis using a lifetime Markov model. The model has been built using Microsoft Excel® and compares the cost-effectiveness of azacitidine with three conventional care regimes in the treatment of myelodysplastic syndrome (MDS):

- Best supportive care (BSC)
- BSC and low-dose chemotherapy (LDC)
- BSC and standard chemotherapy (SDC)

The model is based on the three-year Study AZA-001 data, which are extrapolated to provide a lifetime cohort model. Whilst MDS patients have a short life expectancy and many died within the three year trial period of Study AZA-001, a minority of patients are expected to survive for an extended period beyond this time. Costs are assessed using an NHS and Personal Social Services perspective and cost-effectiveness estimates are expressed in terms of the costs per QALY of azacitidine as an adjunct to standard care (BSC, LDC and SDC).

Alternative analysis timeframes were reported in the sensitivity analyses provided by the manufacturer:

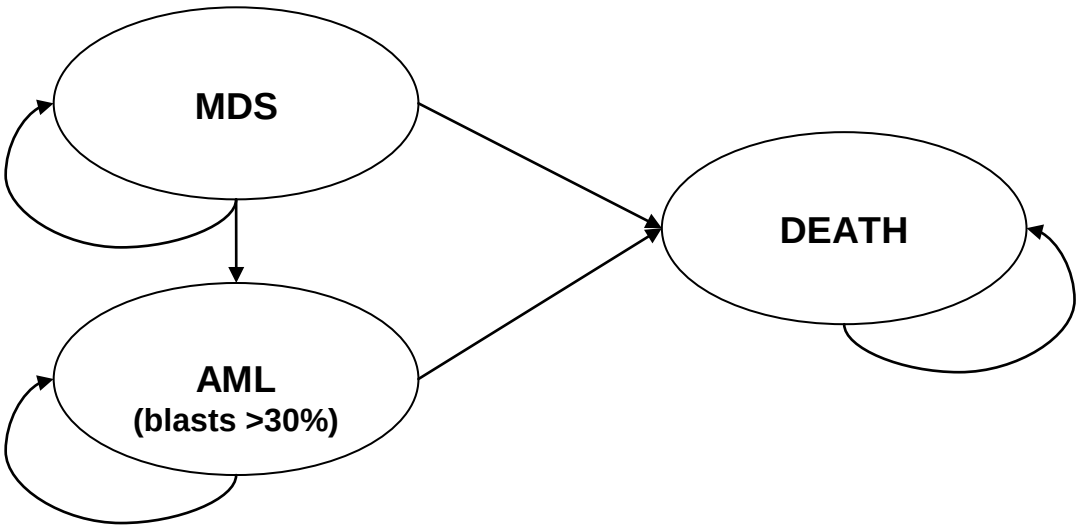
- A three-year time timeframe reflecting the trial period of Study AZA-001
- A year-on-year analysis showing the effect on the incremental cost-effectiveness ratio (ICER) of increasing the model timeframe by one year at a time.

The model is divided into three health states: MDS, AML (acute myeloid syndrome) with blasts >30% and death. All modelled patients enter the model in MDS state on treatment initiation, with death an absorbing state. The model is run for 271 cycles (by which time almost all patients have died), with any additional treatment lifespan ignored.

Patients on active therapy enter the model at the first dose. Patients treated with azacitidine, low-dose chemotherapy or standard-dose chemotherapy are assumed to have been already treated with appropriate pre-medications before entering the

model and these medications have been included in the model. Patients on BSC alone are assumed to enter the model at the same time as they would have entered the model had they been randomised to one of the active therapy arms.

Figure 7.1. Schematic of the health economic model



Patients can die when in either the MDS or the AML health state. It is assumed that a median patient would spend the same amount of time in an AML state, regardless of the treatment received. (The mean time in AML may differ depending on the treatment received.) Once patients progress to the AML state, all patients receive BSC-type treatment, regardless of their previous treatment arm. The transitions between these states are shown in Figure 7.1 of the submission.

Within the original submission, the model encompasses:

- Outcome measures for costs of drugs and medications (Tables 7.10, 7.11 and 7.12 of the submission), monitoring (Table 7.9 of the submission), routine follow-up (Table 7.8 of the submission) and adverse event (Tables 7.13, 7.15 and 7.16 of the submission). Resource utilisation relating to the routine management of patients undergoing treatment is based on expert opinion. It also assumed that there is no wastage within the cost calculations of low-dose

or standard dose chemotherapy, whereas wastage is included in the cost of azacitidine.

- Health outcomes (life-years and quality-adjusted life-years) and incremental cost-effectiveness (Figures 7.8 – 7.10 and Tables 7.17 and 7.19 – 7.28). Utilities for patients treated with azacitidine and BSC are based on mapping European Organisation for Research and Treatment of Cancer scores from Study CALGB 9221 to EQ-5D values using a published algorithm (Tables 7.5, 7.6 and 7.7, and Appendix 6 in the submission).

- A probabilistic sensitivity analysis (PSA) to assess the overall effect of the uncertainty in the model (Figures 7.6 – 7.7c and Appendix 9 of the submission).

As the ERG analysed the manufacturer's model, it became increasingly concerned about the quality of the model provided. These concerns related to a variety of issues beyond the remit of the ERG to correct. Significantly:

- The model contained critical flaws. To the degree that the model was executable, it was only executable because of coding flaws. Where these coding flaws were corrected, the model did not provide data for the case of best supportive care within the assumptions used in the original manufacturer's submission.

- Costs were not discounted within the executable model. The lack of discounting in the executable model significantly affects the cost-effectiveness of adding azacitidine to an existing treatment option.

- The variables and parameter values in the executable model did not correspond to those contained in the submission document.

- In the probabilistic sensitivity analysis, the survival data within the model was coded incorrectly. For both the Log-logistic and Weibull survival curves, this imposed an unrealistic correlation between the variables. Uncertainty in the cost and utility estimates were also characterised incorrectly.

- To the degree that LDC, SDC and BSC are alternative therapies for at least some patients they should be compared together in an incremental analysis. The manufacturer's submission does not do this.

- Many input variables were missing in the model, producing errors (#NAME, #REF). The functionality required to produce some analyses in the manufacturers report had been removed from the executable model. The executable model also appeared to contain several spreadsheets which did not influence the cost-effectiveness results.

- Mortality in the executable model was based on mortality within the three-year AZA-001 Study. Given that the model for this cohort for 26 years. It is reasonable to expect that underlying mortality may differ between, for example, 70 year olds and 96 year olds. Survival is overestimated in the manufacturer's model and particularly at older age groups.

The ERG attempted to repair these errors and reproduce results in the executable model but was hampered by these deficiencies. With regret, the ERG concluded that the model was internally incomplete and could not be fully executed. As such, we were unable to provide results for the AC within the main ERG report.

The ERG requested specific information from the manufacturer, highlighting those pieces of information that would be required in order to correct the model. In response to a request from NICE, the manufacturer provided a text submission of specific pieces of information which were partly sufficient to meet the ERG's concerns. At a later date, the manufacturer also supplied a revised executable model that contained the vast majority of the original flaws. Within the very stringent time constraints, the ERG has attempted to consider this revised model.

The results of this revised economic analysis were summarised in the resubmission text as Table 7.17, reproduced below:

Table 7.17. Summary of base-case cost-effectiveness results

Treatment option	Costs incurred	QALYs gained	Marginal costs	Marginal QALYs gained	Incremental cost per QALY gained
Preselected for BSC					
Azacitidine	£122,035	3.00	£83,677	1.64	**£51,139**
BSC	£38,358	1.36			
Preselected for low-dose chemotherapy					
Azacitidine	£126,061	3.11	£73,826	1.56	**£47,178**
LDC	£52,235	1.55			
Preselected for standard-dose chemotherapy					
Azacitidine	£113,216	2.56	£48,328	1.41	**£34,207**
SDC	£64,888	1.15			

Key: BSC: best supportive care; LDC: low-dose chemotherapy; QALY: quality-adjusted life-year; SDC: standard-dose chemotherapy

According to this latest manufacturers' submission using azacitidine in addition to existing treatment (BSC, LDC, SDC) provides an incremental overall survival gain of around 1.5 years against the comparator arms; across the arms, the greatest QALY gain is observed in best supportive care. Treatment is costly, however, with incremental costs of at least £48,000. These results lead to an ICER of £51,139 per QALY compared with BSC, £47,178 per QALY compared with low-dose chemotherapy and £34,207 per QALY compared with standard-dose chemotherapy.

3.1 End of life

The manufacturer argues that azacitidine fulfils the criteria for the appraisal of end-of-life treatments.

3.2 Natural history

The cost-effectiveness section of the manufacturer's submission does not describe the natural history but this is described elsewhere.

3.3 Treatment of effectiveness within the submission

Treatment effectiveness was largely dealt within the ERG report and the section above (overall survival and time to treatment progression). In the original submission, a log-logistic fit is assumed for modelling base case overall survival and time to AML. A Weibull fit was used in the sensitivity analysis (Tables 7.4a and 7.4b, and Figures 7.5a – 7.5f). In the revised submission, the manufacturers use a log-logistic fit for overall survival. From the original submission (superscript note added):

> Median overall survival was 24.5 months on azacitidine, compared with 15.0 months in the conventional care regimens group (p=0.0001). In a supportive analysis, this survival advantage was observed across all IPSS cytogenetic subgroups, in patients with –7/del(7q) and in elderly patients with AML. The overall survival gain was observed despite relatively low response rates. Analysis suggests that achievement of complete remission is not essential to improve survival. Partial remission and haematological improvement were also associated with survival benefit.
>
> The reduction in risk of death on azacitidine compared with CCR was 42% (p=0.0002). At two years, the proportion of patients surviving was approximately twice as high in the azacitidine group as in the CCR group (50.8% versus 26.2%; p<0.0001). The median time to transformation to AML was also greater in the azacitidine group (17.8 versus 11.5 months; p<0.0001). In summary, azacitidine significantly lengthens overall survival in patients with higher-risk disease (IPSS categories intermediate-2 and high)[a] (…)
>
> Patients have an increased mortality risk when in the AML health state. It is assumed that the risk of mortality in AML is independent of the patient's previous treatment arm and the pooled mortality rate is used across all treatment arms. This results in an AML mortality rate of 0.135 per cycle. This rate is applied to all patients in the AML health state. The rate of mortality of patients in the MDS health state is calculated by adjusting the rate so that the overall survival rate is maintained.

[a] See Figures 6.5 and 6.6, and Tables 6.7-6.9 of the manufacturer's submission.

Comments:

- There are inconsistencies within the executable model regarding data columns registering events as “censored” or “dead”. However, the survival inputs appear correct and whilst the impact of these issues is unclear, it is possible that the economic analysis is correct. Further investigation of the issue was not possible due to a lack of clarity in the submission.

- It is not clear whether Figures 7.5a to 7.5f of the manufacturer’s submission use Weibull- or log-logistic-fitted survival estimates. MS figure 7.5c appears in error since the overall survival and TTP curves cross over, which would indicate that some patients progress to AML after death.

- The *original* submission uses a median time in AML of 3.65 cycles. TTP is modelled by fixing one model parameter and varying the other to reduce median overall survival (but not mean overall survival) by 3.65 cycles. However, when modelling mortality in AML the manufacturer’s claim a fixed rate of 0.135/cycle is employed, which imposes an exponential distribution on AML survival with a median of 5.1 cycles. This would appear to extend survival beyond that which can be logically justified. However, this mortality rate (0.135 per cycle) rarely applies within the executable model. Overall mortality is calculated and the 0.135 per cycle figure is used to allocate this typically lower mortality between the AML and MDS states. Where total estimated mortality is lower than the expect AML mortality, AML mortality is reduced to match the survival data, with MDS mortality reduced to zero. This is the case in the vast majority of cycles.

- Mortality in the executable model was based on mortality within the three-year AZA-001 Study. The AZA-001 cohort faced distinct types of mortality risk from MDS/AML and from unrelated causes. Since MDS is experienced at baseline, the model may be appropriately sensitive to these risks (subject to the issues above). However, the model is not sufficiently sensitive to mortality from these unrelated causes, which differ by both age and gender. The model is run for 26 years and it is clear that these underlying mortality risks will differ

across this period. Overall survival, especially longer term survival, is likely to be seriously overestimated in the manufacturer's model.

Comment on the choice of log-logistic fits for the base-case analysis

When log-logistic fits to overall survival are extrapolated to 312 cycles (29.9 years) about 4 to 5% of azacitidine-treated patients remain alive. Adding to the mean age for each subgroup indicates that in this model 4 to 5% of azacitidine MDS patients survive into their tenth decade. This considerable azacitidine-dependent improvement to the generally accepted prognosis for MDS patients is not observed when Weibull distributions are fitted. Clinical opinion may consider that the Weibull extrapolation has greater face validity and is more suitable for the base case economic analysis. Weibull and log-logistic fits to observed overall survival for the three subgroups are shown in section 2.2.

In response to an ERG request for clarification the manufacturer provided Kaplan-Meier analysis of overall survival for those azacitidine-treated patients in the CALGB9221 study with similar disease severity to patients in the AZA-001 RCT (see section 2.2). All patients were followed to death and all were dead by 6.3 years (median survival 22.5 months similar to the 24.1 months in the AZA-001 trial). This result supports the Weibull fit to overall survival in AZA-001 rather than the log-logistic fit used in the base case.

Comment on the model strategy for estimation of time to progression to AML

To avoid using observed TTP the model employed a pooled estimate for median time in AML (3.65 cycles) and modelled TTP using a back calculation procedure: A] by deploying the relationship "Median time to progression to AML = median time for overall survival - median time in AML"; B] by assuming a common median time in AML of 3.65 cycles for all patients; C] by applying a log-logistic distribution for TTP that was defined by the shape parameter of the log-logistic fit to overall survival.

This strategy does not alleviate the problem of poor assessment of progression and is unlikely to be reliable for the same reasons that the observed TTP was considered unreliable; that is because a good pooled estimate of the median time in AML depends on reliable identification of those patients who progressed and of the time of their progression. Neither of these conditions was fulfilled because bone marrow assessments were [redacted] (see ERG clinical evidence section)

Three suggestions were proposed to justify this approach; these are listed below together with ERG comments:

1] Clinical advice saw no reason that time in AML should differ between treatment arms or subgroups.

> *Comment:* there were no details of how the clinical advice was obtained or how many clinicians were involved. There was no comparison of the fixed time of 3.65 cycles in AML with the observed time in AML for the different patient groups. There was no indication of the uncertainty associated with the estimate of median time in AML.

2] There were so many censorings in the observed TTP analysis that no reliable estimate could be obtained for subgroups (e.g. SDC subgroup, MS fig 7.4, Page 81).

> *Comment*: the estimate of TTP based on the observed progressions in the AZA-001 trial was unreliable because of poor ascertainment of progression. In particular, many scheduled bone marrow assessments were [redacted] for the [redacted].

3] The modelling of TTP would result in TTP not driving the cost-effectiveness analysis.

> *Comment*: There is no reason to expect that modelling will diminish the influence of TTP on the cost-effectiveness results. The model defines two states for live patients, MDS or AML, and each is associated with different costs and different QoL utility. A prolonged TTP means fewer patients in AML and more in MDS (and vice versa). When TTP differs between compared groups (as in the MS) the differential partition of live patients between the two states inevitably has some influence on cost-effectiveness results.

3.4 Health related quality of life

Health effects are expressed in both life years and QALYs gained. The estimates in the manufacturer's model suggest that those given azacitidine may have improved quality of life. Given few data on the utility MDS patients and no Study AZA-001 quality of life data, this estimate is based on mapping clinical scores to EQ-5D data.

From the original submission:

> Utilities for patients treated with azacitidine and BSC are based on mapping European Organisation for Research and Treatment of Cancer (EORTC) scores from Study CALGB 9221 to EQ-5D values using a published algorithm by McKenzie et al. QoL scores for low-dose and standard-dose chemotherapy reported by the SF-12 QoL instrument in Sekeres et al. are mapped to EQ-5D values by applying a published algorithm by Gray et al.
>
> Full details of this mapping process are given in Appendix 5.
>
> The utility analysis results specifically show that there is a treatment effect on the QoL of patients in the MDS health state. Patients treated with azacitidine have a better QoL than those treated with BSC, and this difference increases with increasing length of treatment. The utility scores seen at 182 days are assumed to remain constant for the remainder of the patient's time in the MDS health state[2].
>
> There are no utility values available in the literature for patients that are in the AML (blasts >30%) health state. In the base case, this value is assumed to equal the baseline MDS utility score of 0.67.The effect on the ICER of varying this figure is examined in a sensitivity analysis, the results of which are shown in Figure 7.10.The analysis shows that varying the utility score of the AML (blasts >30%) health state has minimal effect on the ICER (figure 7.10).

[2] See table 7.6 of the manufacturer's submission

AEs do not have AE-specific utility detriments modelled. This was to prevent any potential double-counting of the health effect due to the capture of longitudinal utility data.

Comments:

- The utility for the AML state is assumed equal to the baseline utility for MDS. As this utility includes the states immediately before death, a utility of 0.670 may be considered high.

3.4.1 Comments on mapping used in the model

Although use of mapping algorithms is permitted by the NICE methods guidance in economic evaluations of treatments, the methodology behind mapping should still be considered immature. Any mapped estimate used in such an evaluation must be interpreted with caution and with close consideration of the estimation methods used. In this section we describe the processes used for mapping in the manufacturers submission (see Appendix 5 of the submission for full details) and appraise the method used for the bases case analysis and the original paper on which the algorithm was based.

The manufacturer's provide, in Appendix 5, a report produced on their behalf describing the process by which utility values were derived for this submission. This work was undertaken in three phases, each phase testing different approaches to estimation.

In Phase 1, mean EQ-5D utility estimates were estimated for azacitidine and best supportive care. The source data for these estimates was taken from Kornblith et al (2002) (section 7.2.8.3 of the original submission). The algorithm used to convert this data to EQ-5D scores was based on a published mapping coefficients derived by McKenzie and van der Pol (2009). Kornblith et al do not report mean values for all functionings in the EORTC QLQ-30 and therefore only a partial mapping can take place, using those functionings that were reported. The authors of Appendix 5 argue that the most relevant data were reported in Kornblith and that therefore the lack of all functionings for estimating EQ-5D scores should make little difference. This

assumption is not (and cannot) be empirically tested but may introduce significant uncertainty into the mapped estimates.

Phase 2 used the same mapping coefficient estimated by McKenzie and van der Pol as in Phase 1. However, this approach differed in that it estimated EQ-5D scores using the individual patient level data used in Kornblith et al. This overcomes the key failing of the Phase 1 approach, in that all functionings of the EORTC QLQ-30 measure are accounted for, potentially reducing uncertainty in the result.

The approach used in Phase 3 differed significantly than that used in Phases 1 and 2. In Phase 3 the authors use three approaches. The first was to use the individual data set from a publication by Sekeres at al (2004) and SF-6D utility values. The SF-6D values were derived from SF-12 data in the Sekeres paper based on an algorithm published by Brazier and Roberts (2004). The second approach was to map EQ-5D data based on the Physical Component Score and Mental Component Score of the SF-12 and using an algorithm published by Franks (2004). The third report was to use the approach published by Gray et al (2006).

The approach used in the manufacturers submission is based on the Phase 2 work as presented in detail in Appendix 5. The estimates reported in Table 10.9 of Appendix 9 of the manufacturer's original submission (pare reproduced below.

Variable	Mean value	SD	Distribution
Utilities			
Azacitidine day 0	0.67	0.22	Beta(2.39,1.18)
Azacitidine day 50	0.70	0.20	Beta(2.98,1.28)
Azacitidine day 106	0.74	0.20	Beta(2.82,0.99)
Azacitidine day 182	0.80	0.21	Beta(2.1,0.53)
BSC day 0	0.67	0.22	Beta(2.39,1.18)
BSC day 50	0.69	0.20	Beta(3,1.35)
BSC day 106	0.68	0.22	Beta(2.38,1.12)
BSC day 182	0.72	0.22	Beta(2.28,0.89)
SDC day 0	0.66	0.13	Beta(8.1,4.17)
SDC day 14	0.61	0.1	Beta(13.9,8.89)
SDC day 42	0.66	0.1	Beta(14.15,7.29)
SDC day 70	0.69	0.12	Beta(9.56,4.29)
SDC day 98	0.72	0.16	Beta(4.95,1.93)
SDC day 182	0.74	0.18	Beta(3.65,1.28)
SDC day 365	0.83	0.1	Beta(10.88,2.23)
LDC day 0	0.67	0.08	Beta(22.48,11.07)
LDC day 14	0.70	0.09	Beta(17.45,7.48)
LDC day 42	0.71	0.15	Beta(5.79,2.36)

Variable	Mean value	SD	Distribution
LDC day 70	0.72	0.13	Beta(7.87,3.06)
LDC day 98	0.70	0.06	Beta(40.13,17.2)
LDC day 182	0.85	0.08	Beta(16.08,2.84)
LDC day 365	0.67	0.22	Beta(2.39,1.18)

It is the case however that there are a two key caveats highlighted in the paper by McKenzie and van der Pol regarding the reliability of the mapping coefficients. These must be born in mind when considering whether or not these mapped utility estimates can be considered reliable.

The first is that in some cases a very small number of patients in the EORTC QLQ-30 data set and EQ-5D data set reported very severe problems. This can lead to problems with developing a consistent model for the data. Mapping algorithms are derived using a variety of regression techniques. Where there are small sample sizes for certain domains or functionings, regression methods will not be able to generate a reliable estimate of the predictive nature of moving from one instrument to the other. In that case, we may find that the algorithm is more reliable in predicting the utility values for patients in less severe states but less reliable for predicting them for patients in more severe states, leading to bias in the results

The second issue with the algorithm from McKenzie and van der Pol, is that, in the words of the authors, further research is needed "…before the application of the model should become a recommended approach for converting the EORTC QLQ-30 data into EQ-5D values". A key reason given for this conclusion include questions over the generalisability of the model to other patient groups (the McKenzie and van der Pol model was developed using data from patients with esophageal cancer). While the patients eligible for azacitidine are of a similar age group to the patients in the study, the underlying conditions and co-morbidities are potentially very different. Taken in conjunction with the lack of confidence one can have in the mapping functions predictive nature for patients with severe answers in different domains or functionings, the mapping algorithm used in this study should be treated with caution.

3.5 Resources and costs

Only 6% (23/358) of those enrolled in Study AZA-001 were from UK centres. The resource implications of azacitidine were obtained from a survey of UK haematologists, as the manufacturer deemed that expert opinion would be necessary to provide healthcare resource data relevant to the NHS. Healthcare resource use for the management of MDS and AML (blasts > 30%) were obtained using a structured questionnaire that assessed administration, testing and follow-up costs.

The costs associated with low-dose chemotherapy and standard-dose chemotherapy were based on the Study AZA-001 protocol dosage regimens. Costs relating to adverse events (Grades 3 and 4) were taken from NHS Reference Costs by speciality, with rates derived from expert opinion. For disease-related complications the manufacturers use Study AZA-001 costs.

Comments:

- It is unclear on what basis Study AZA-001 was considered appropriate to provide some resource information but not others.

- A footnote in Table 7.13 of the submission informs that costs related to adverse events such as pyrexia, pneumonia and sepsis are assumed to be treated the same as febrile neutropenia, as advised by expert haematologists. Given that adverse events are a major driver of uncertainty, a more explicit discussion on this assumption appears warranted. Further discussion on this assumption would have been welcomed.

3.5.1 Discounting

The submission claims that a 3.5% annual discount rate was used for costs and health benefits. This would be in line with the NICE reference case.

Comments:

- Within the original executable model, the manufacturers neglected to discount costs, which increased the incremental cost-effectiveness ratios for azacitidine.

- In the revised executable model, cost and QALY discount rates are calculated according to “years” of either 10 or 11 35-day cycles (as appropriate). A cycle-specific discount rate for costs and outcomes of 3.5% × 35/365 = 0.46% would be more appropriate.

3.6 Sensitivity analyses

3.6.1 One-way sensitivity analyses

Both the manufacturer’s original and revised submissions provide Tornado diagrams to represent one-way sensitivity of results. Unfortunately, neither of the models provided was fully executable as neither allowed Tornado diagrams to be constructed. The presence of such diagrams in the text submissions suggests that this analysis was feasible. The manufacturers claim that the main source of uncertainty in the model appears as the application of AEs but this could not be confirmed by the ERG. Within the original submission, the manufacturer considers scenarios in which:

- The annualised AE rates for azacitidine, BSC, low-dose chemotherapy and standard-dose chemotherapy are applied in each cycle in which the patient is on treatment. Once patients move off treatment, the annualised AE rate for BSC is used.

- The annualised AE rates for azacitidine, BSC, low-dose chemotherapy and standard-dose chemotherapy are applied to patients throughout their time in MDS.

The same scenarios are considered in the revised analysis. Results regarding conventional care regimen as a comparator treatment arm are also provided. The results of this revised sensitivity analysis were summarised in the resubmission text as Table 7.21, reproduced below:

Table 7.21. Sensitivity analysis of the methodology of applying AE rates

Comparator treatment arm	AE methodology		
	Base case	Annualised rate on treatment (1)	Annualised rate in MDS (2)
Conventional care regimen	£47,687	£46,635	£44,572
Best supportive care	£51,139	£49,853	£50,189
Low-dose chemotherapy	£47,178	£46,006	£46,378
Standard-dose chemotherapy	£34,207	£34,369	£17,502

Key: AE: adverse event; MDS: myelodysplastic syndrome

3.6.2 Probabilistic sensitivity analyses

The manufacturers also provide a probabilistic sensitivity analysis. In the original submission, the reporting of this analysis (Appendix 9) was inconsistent with the executable model with respect to both the variables contained in the analysis and the values of these variables. The ERG had concerns about elements of the probabilistic sensitivity analyses in both versions of the executable model. Whilst some issues were made clarified following requests from NICE, others remain and seriously impact upon the credibility of the manufacturer's estimates of uncertainty. These residual concerns relate to:

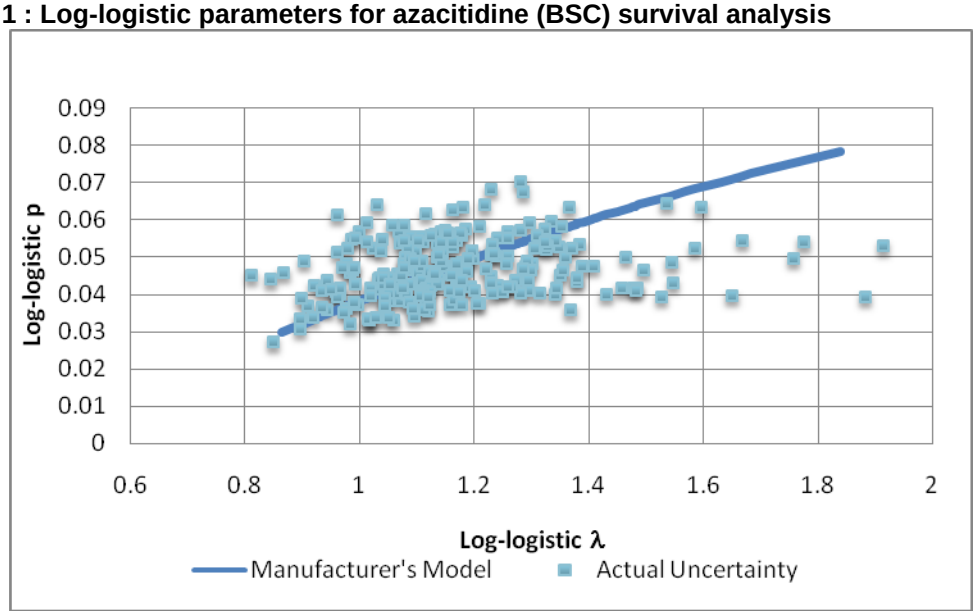
- The use of survival analysis
- The estimation of uncertainties in utility data
- The estimation of costs

1) The use of survival analyses within the PSA

When considering each time to progression in AML, the probabilistic sensitivity analysis uses one random number generation to estimate values for two parameters. This implicitly assumes a perfect relationship between the two parameters in the Log-logistic and Weibull regressions. As coded in both the original and revised manufacturer's models, the relationship between the Log-logistic parameters assumes that both parameters are related in an approximately linear way (the solid

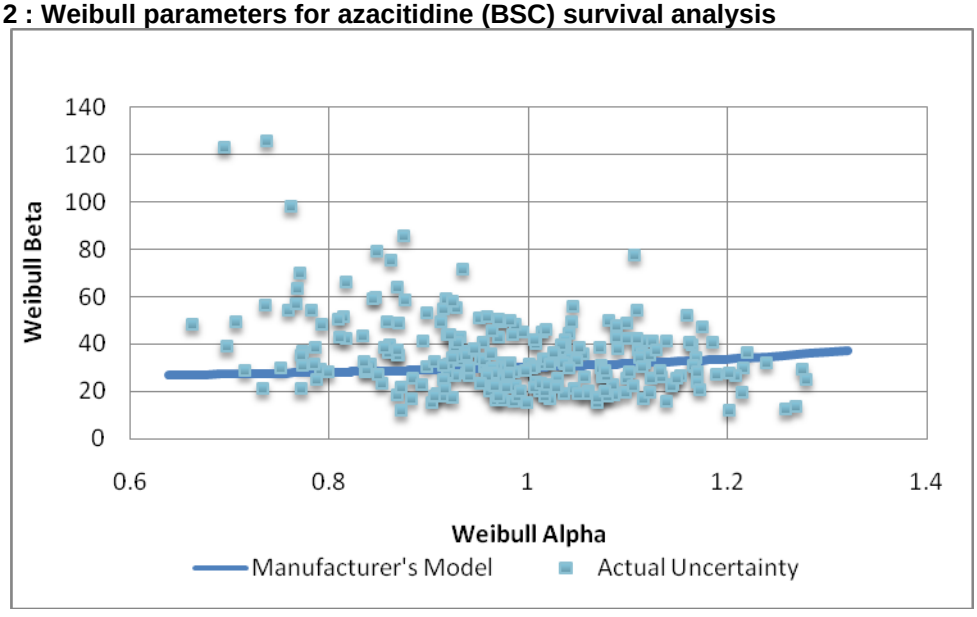
line below). The clarifications provided by the manufacturer reveals that the survival analysis does not support this assumption, with a relatively weak correlation between the parameters (between 0.08 and 0.38). In Figure 1 below, the Log-logistic parameters are analysed for the azacitidine (BSC) case. There is substantially more variation in these parameters than acknowledged by the manufacturer in its models.

Figure 1 : Log-logistic parameters for azacitidine (BSC) survival analysis



The equivalent analysis for the Weibull parameters also reveals flaws (Figure 2). In the survival analysis (the square points), the parameters exhibit a strong *negative* correlation (typically between -0.95 and -1). As coded, the model assumes a positive correlation.

Figure 2 : Weibull parameters for azacitidine (BSC) survival analysis



The ERG believes that the manufacturer's model fundamentally misrepresents uncertainty in this regard. This is the most serious of the issues with the probabilistic sensitivity analysis, and means that little confidence can be placed in the

manufacturer's PSA results even if the underlying deterministic model is deemed valid by the AC.

2) The estimation of uncertainties in utility data

Within the PSA, estimates for the utility parameters were obtained assuming that the only relevant source of uncertainty related to the utility of one treatment over another. Within a treatment, the model assumes no uncertainty for the utility observed in one period over the utility observed in another. Beta distributions are estimated for each utility observation using mean and standard deviations. For each treatment, a random number is generated to represent a point to select in the cumulative beta distribution, with the same random number used for all observations within a treatment. In this way, there is a relationship assumed in the rankings of each state.

This approach would be more justifiable if we had reason to believe that the rankings across time reflected a logical ordering. If we were considering "poor", "good" and "excellent" health, we might expect that "poor" to always be worse than "good", and "good" to be worse than "excellent", and the methodology used would always provide such results. It is far from clear that Day 0 utilities should be any better or worse than Day 42 utilities, or Day 42 to be better or worse than Day 98, and so on. But this is exactly what the manufacturer's approach suggests.

The rigorous approach would be to consider the correlations between the observations and to incorporate this within the analysis. As the manufacturer could not provide this data, the ERG consider that the appropriate response is to treat each utility observation as an independent variable and incorporate these into the model using a different random number generation for each utility in the model.

The ERG believes that the manufacturer's model underestimates the level of uncertainty in this case.

3) The estimation of uncertainties in cost data

Within the estimation of treatment costs, the manufacturer generates two random numbers for each cost item. It appears that the first number is compared to the proportion of clinicians using each cost item, and the item is "used" if the random number is sufficiently high. Where "used", the model estimates a cost from the distribution of costs amongst those clinicians who indicated that they use the item (using the second random number). All the random numbers used were independently drawn.

This approach appears to be valid in estimating uncertainties in the cost of treatment of a single clinician. However, it is much less clear that it gives an accurate assessment across all clinicians. From the clinician survey it is clear that there is substantial heterogeneity in the clinical approach to MDS, and as such we would expect a variety of approaches to be observed in reality. In this sense, the approach may overestimate uncertainty by considering extremes in individual clinical practice rather than the general approach within the clinical community.

Because the selection of items within a scenario is independent, it is also possible that scenarios may include combinations of resources that lack face validity. For example, the model assumes that for Best Supportive Care, 59% of scenarios use Senior Nurses and in 29% Junior Doctors are used. It might be feasible to expect that these resources are substitutes, in that most of the time either a Senior Nurse or a Junior Doctor is used. In the model, around 18% of scenarios (59% x 29%) will use both groups in Best Supportive Care and 28% of scenarios (41% x 71%) will use neither. If this criticism is valid, then the manufacturer's methods may also overestimate uncertainties in a different way.

The ERG considers that the rigorous approach would incorporate this heterogeneity, and the correct method to use would be to bootstrap to find estimates of mean usage (rather than the usage of a single clinician) from the results of the clinical survey. It was not possible for the ERG to do this. The ERG believes that the manufacturer's model is likely to overestimate the underlying level of variation in cost estimates.

3.7 Model validation

The ERG observed several issues with the executable model provided by the manufacturers which seriously limited the degree to which results could be verified. Even after the resubmission of the manufacturer's model, the ERG considers that serious problems remain.

The ERG was not able to validate all presented model results given time constraints. The ERG was not able to fully validate the manufacturer's revised model given time constraints. To the degree that the ERG considers the model flawed, the Group has attempted to rectify several of the manufacturer's major errors (in addition to other minor errors not provided in detail here for reasons of brevity). The ERG considers that there are sufficient grounds to suggest that the AC consider the appropriateness of both the model's face validity and assumptions.

3.8 Comment on validity of results presented with reference to methodology used

The errors found by the ERG in the original model prevented the model from running as stated. The revised manufacturer's submission has greater validity inasmuch as it now runs (and includes discounting). Finally, in their initial assumptions, the manufacturers do not consider other chemotherapy agents in their submission outside of these broad categories whilst available. They suggested that there was no evidence that these regimes omitted are superior. Nevertheless, the ERG underlines that if cost-effectiveness differs amongst the agents on alternative "generic" chemotherapy, the picture might be different.

3.9 Summary of uncertainties and issues

The ERG has no confidence in the deterministic results of the model given that: 1) the death rate from AML is typically far below the 0.135 assumed in the manufacturer's model, 2) the model often does not consider death within MDS states, 3) age related mortality is not included, 4) there serious issues surrounding the choice of functional form to characterise survival data, and 5) even if the

functional form is correct, we have little confidence in the method used to calculate time to progression.

The ERG has no confidence in the probabilistic sensitivity analysis provided and could not verify the one-way sensitivity analysis. The key issues here are 1) the mischaracterisation of uncertainty in the survival analysis, and 2) strong assumptions used in characterising uncertainty in utilities. Whilst we have concerns about the way cost uncertainties are characterised in the model, we are not able to modify this.

Finally, the results provided by the manufacturer do not necessarily correspond to the clinical question, in which the choice is probably between (in order of treatment intensity): BSC, BSC + azacitidine, LDC, LDC + azacitidine, SDC, SDC + azacitidine. Whilst not all these options are available for all patients, more than two options are available to many. Whether or not treatment is cost effective will depend on other factors, including the age of the patient and the likely severity of side effects from chemotherapy. This would appear to be a key part of the clinical problem but is not addressed within the model.

4 Additional work undertaken by the ERG

4.1 Work undertaken by the ERG

The ERG has serious concerns about the revised model provided by the manufacturer. Some of these concerns were partly addressed by subsequent modification of the manufacturer's model. These included many minor changes which decreased the number of cells containing error terms, and others where obvious errors were corrected (including where the PSA was unable to incorporate sources uncertainty because cells used only deterministic values).

The baseline model for the ERG modifications used the same settings as the baseline manufacturer model. Here, AML transitions were based on a modified survival analysis, as above. The revised manufacturer model used different settings than the original manufacturer's model. In the revised manufacturer's model, the rate of AML was equal to the rate of death six cycles into the future. This effectively

assumes that no people die in the MDS state, which appears to be questionable given the clinical information provided by the manufacturer.

For the manufacturer, the change of assumptions provided a model giving similar results, since the coding errors in the previous model had a similar effect. The manufacturer does not detail how the underlying clinical knowledge about MDS has changed to justify a departure from their initial assumptions.

The manufacturer was suggested that correcting flaws in the model does not affect results. However, this claim only appears to hold when they make fundamental changes to the assumptions underlying the model. Whilst the broken links provided did not affect the results used in the *revised* evidence submission, this was only because the revised evidence submission used changed assumptions that mimic those erroneously used in the flawed model.

Relative to the *revised* model using the *original* settings, the substantive changes in the ERG analysis were that:

1) An incremental analysis is computed in which cost-effectiveness frontiers are calculated across all six options for those receiving SDC, across BSC-based and LDC-based options for those receiving LDC, and BSC is again compared to BSC + azacitidine. This analysis may provide a clearer account of the clinical issues (with caveats attached with respect to how far we can compare across groups).

2) Discount rates were modified using a per-cycle rate rather than an estimate of a yearly discount rate.

3) Given the mischaracterisation of uncertainty in the survival data, the covariance between parameters was included in the model using two parameters and a Cholesky decomposition. The manufacturer did not provide all relevant covariances in the model, and where these were not provided the ERG assumed a correlation of 0.20 and -0.95 for the log-logistic and Weibull regressions. These figures are generally in line with those provided by the manufacturer. This substitution includes one case in which the provided figures for log-logistic and Weibull variance/covariance were identical (LDC + azacitidine), suggesting a transcription error from the manufacturer.

4) Within the PSA, the characterisation of uncertainty in the utility parameters was changed as indicated above.

5) Cost-Effectiveness Frontiers are calculated for those deemed able to receive SDC and LDC, and a Cost-Effectiveness Acceptability Curve is calculated for those receiving BSC.

Of these changes, 1-2 relate to the deterministic model and 3-5 relate primarily to the probabilistic sensitivity analysis.

As a purely indicative analysis, the ERG has also considered the impact of incorporating age-specific mortalities. Assuming that the cohort is made of 70 year old males, life tables from England and Wales were used to obtain yearly mortality rates. These were linearly interpolated to provide a per cycle mortality figure. Data for ages 70-72 were used to obtain within-study standard mortality rates. For Cycle 32 onwards (approximately Year 4 onwards), mortality rate was modified by the age-specific mortality minus the baseline mortality from the first three years. Note that the importance of age-specific mortality increases with the numbers surviving to older ages. Given the problems identified with the AML mortality rate in the model, the major mortality issue is likely to be the fact that survival in AML is exaggerated in the model.

Given that the ERG has imposed a correlation of 0.20 for cases where no data was provided by the manufacturer, the Group also considers a case in which the correlation is removed. In these cases, survival parameters where no information was provided are estimated to be independent random variables in the probabilistic sensitivity analysis.

4.2 Work not undertaken by the ERG

Given time and resource constraints, we were unable to consider all elements of the manufacturer's submission. The amount of modification to the model conducted by the ERG is beyond the scope normally provided within an STA process and reflects the degree of concern with the model. The analysis from the ERG has been informed by what has been possible in the very short time we have had a near-functional

model. The revised economic submission has been analysed with many of the remaining errors resolved in order to provide a model that allows the AC to consider whether or not the submission ultimately has face validity.

- The ERG revised model does not correct for the underestimate in AML mortality. This is likely to be the key in driving the degree of benefit provided by azacitidine, and the values provided in this model are highly likely to overstate marginal benefits.

- Uncertainty relating to the time spent in AML. This is a key assumption in the manufacturer's model and would ideally have been incorporated into the probabilistic sensitivity analysis. With additional time, it may have been possible to consider this source of uncertainty in more depth.

- The analysis showing sensitivity to the model's time horizon would have been useful to replicate but time constraints did not allow this.

- The ERG revised model does not include an exponential survival analysis, although again this may improve the model.

- The CCR versus azacitidine comparison has not been considered, as the manufacturers acknowledge that methodology fails to meet necessary standards.

- The incorporation of age-specific mortality figures were considered for one age, and for males only, and the ERG does not place confidence in the analysis. A fuller analysis would involve more robust methods, and consider a much wider range of cases.

- The impact of alternative assumptions regarding adverse events could not be explored by the ERG.

- End of life issues are outside the remit of the ERG and thus are not considered.

- Whilst the manufacturer argues that a [redacted] may allow azathioprine to be cost-effective , this functionality did not appear to be included in the model. As such, it is not possible to consider what impact this would have in a more robust analysis.

5 Baseline analysis of ERG model

The results of this ERG base case analysis were summarised in Table 1 below:

Table 1. Summary of base-case cost-effectiveness results: deterministic results

Treatment option	Outcomes		Marginal Differences		ICER (adjunct therapy)
	Costs	QALYs	Costs	QALYs	
Azacitidine BSC	£120,007	2.90	£81,694	1.57	**£51,973**
BSC	£38,313	1.33			
Azacitidine LDC	£123,491	3.01	£72,414	1.48	**£48,766**
LDC	£51,077	1.52			
Azacitidine SDC	£110,472	2.47	£46,503	1.35	**£34,525**
SDC	£63,968	1.12			

Key: BSC: best supportive care; LDC: low-dose chemotherapy; SDC: standard-dose chemotherapy.

Overall, the figures provided here are similar in size to those provided in by the revised manufacturer's model. In all three comparisons versus standard care, azacitidine is costly and provides health gain. In none of the three cases does adding azacitidine provide health gain below a £30,000 per QALY figure.

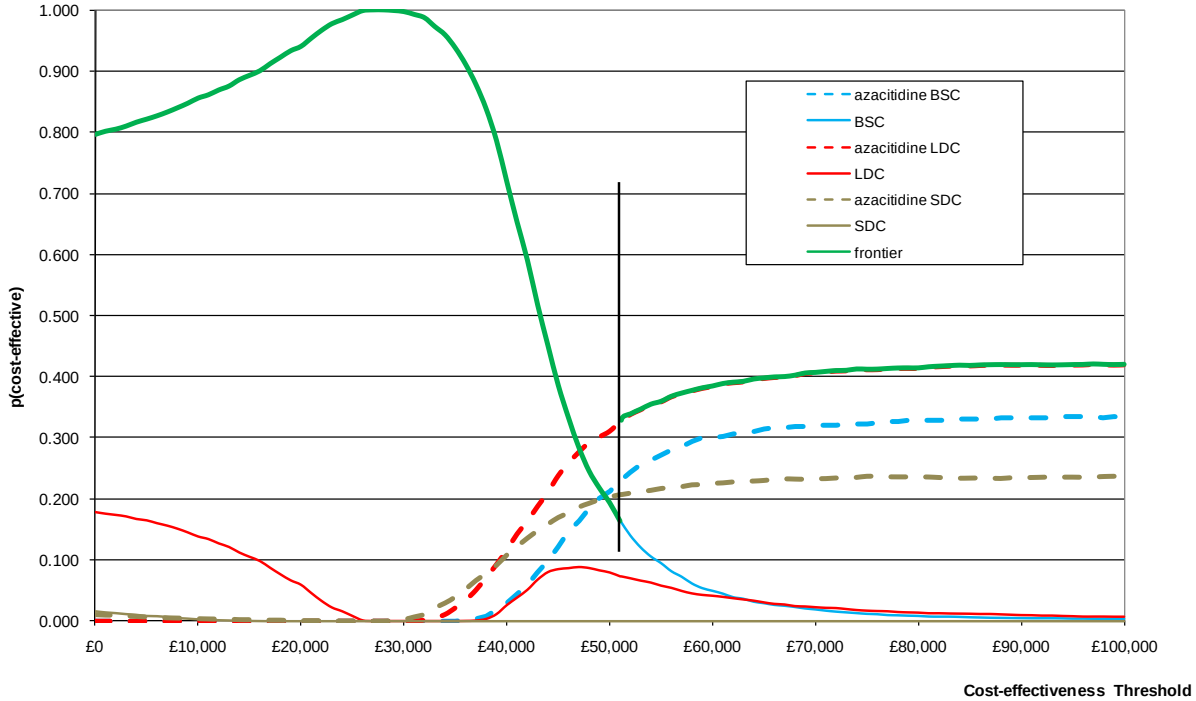
The health gain provided in low dose chemotherapy (LDC) is greater than the health gain provided in standard dose chemotherapy (SDC), at a lower total cost. Within the manufacturer's submission it was suggested that both groups receiving chemotherapy tend to be at high risk, with those in SDC tending to be younger. If so, we might expect those in SDC to have a better prognosis than those receiving LDC. If this is the case, then the manufacturer's estimates Study AZA-001 data suggests that SDC provides for extremely poor outcomes. Either 1) the model lacks face validity, or 2) the model has validity and outcomes must be compared across treatment groups. In the former case, the model based on Study AZA-001 data should be ignored as it does not provide useful information. In the latter case, a more sophisticated analysis is necessary.

For those receiving SDC, LDC would appear to be a treatment option, as is BSC. For those receiving LDC, BSC is a treatment option but it is not clear that SDC would be. For the three groups (SDC, LDC, BSC), the ERG analysis considers 1) all six

treatment options, 2) LDC/BSC, with or without azacitidine, 3) BSC/BSC plus azacitidine.

For the first case, Figure 3 provides a cost-effectiveness frontier. For each value of the cost-effectiveness threshold, this (green) line identifies the outcome with the highest net benefit. Where all six treatment options are available, it appears that the most cost-effective option is Best Supportive Care for all values of the threshold up to £51,135 per QALY. Above £51,135 per QALY, low dose chemotherapy and azacitidine becomes the most cost-effective option. No SDC option is cost-effective. At £30,000 per QALY, the model suggests a 100% chance that BSC is cost-effective. By contrast, SDC is never cost-effective above £17,000 per QALY.

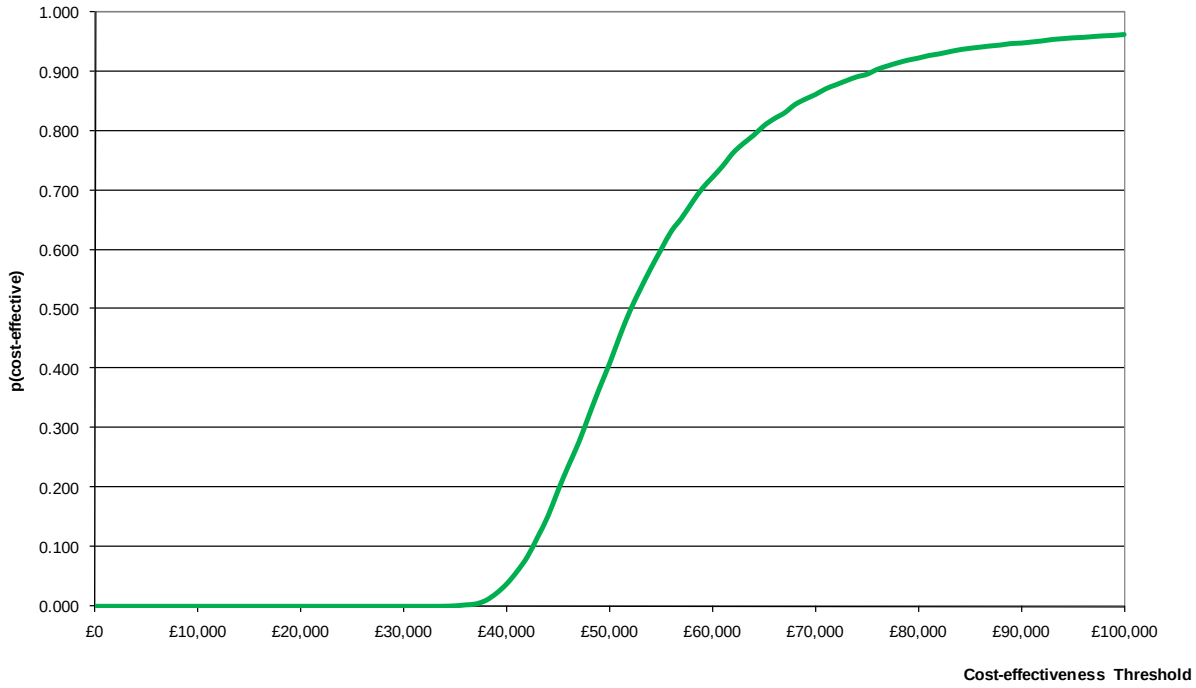
Figure 3: Cost-effectiveness Frontier: SDC options



The only cost-effective options here involve low dose chemotherapy or best supportive care. The same options are optimal over the same cost-effectiveness ranges for those who currently receive low dose chemotherapy.

For those currently receiving BSC, the cost-effective outcome at £30,000 per QALY is clearly to retain BSC rather than BSC + azacitidine. At the ICER for BSC + azacitidine (£51,973 per QALY), 49.8% of model runs suggest that adding azacitidine would improve cost-effectiveness.

Figure 4: **Cost-effectiveness Acceptability Curve: BSC + azacitidine**



The results of this model appear to be reasonably robust with respect to the assumptions used by the ERG when assessing its sensitivity. Table 2 below gives the optimal choices and ranges for the alternative analyses conducted for both the SDC group (where all treatment choices are available) and the LDC group (where the LDC and BSC treatment choices are available). The choice of optimal treatment is largely unaffected by the changes in assumptions below except where the Weibull survival analyses are used instead of the log-logistic results. Here, azacitidine SDC provides the most net benefit across the range where azacitidine LDC is typically optimal in the other sensitivity tests.

Table 2. Summary of model modifications: SDC and LDC groups

Modification	Group	Deterministic ICER (per QALY, as adjunct)	Cost-Effectiveness Frontier	
BASELINE	SDC	£34,525	BSC: λ < £51,135	azacitidine LDC: λ > £51,135
	LDC	£48,766	BSC: λ < £51,135	azacitidine LDC: λ > £51,135
Weibull Survival Analysis	SDC	£44,627	BSC: λ < £53,061	azacitidine SDC: λ > £53,061
	LDC	£64,792	BSC: λ < £54,174	azacitidine LDC: λ > £54,174
Age and gender specific mortality (70 year old male)	SDC	£35,723	BSC: λ < £55,407	azacitidine LDC: λ > £55,407
	LDC	£52,613	BSC: λ < £55,407	azacitidine LDC: λ > £55,407
Without assumed correlations*	SDC	£34,525	BSC: λ < £52,644	azacitidine LDC: λ > £52,644
	LDC	£48,766	BSC: λ < £52,644	azacitidine LDC: λ > £52,644

*Affects PSA only.

All three of the subsequent analyses suggest a larger range over which BSC is optimal. All four analyses suggest that the critical point lies between £51,000 and £56,000 per QALY. The deterministic ICER providing the cost per QALY of each therapy as an adjunct to a basic therapy (SDC, LDC) are also relatively consistent across groups. The ERG emphasises that the underestimate of AML mortality and general non-inclusion of MDS mortality is likely to overestimate the benefits of azacitidine.

For the choice of BSC versus BSC plus azacitidine, the results are also reasonably robust to changes in model assumptions. Against a deterministic ICER of £51,973 per QALY, the Weibull analysis suggests a higher figure at £67,243 per QALY, with the addition of age specific mortality having a smaller impact (£55,644 per QALY). The without-correlation case trivially has the same deterministic ICER as this assumption only affects the probabilistic sensitivity analysis.

Figures for all the cost-effectiveness frontiers and cost-effectiveness acceptability curves referred to here are provided in Appendix 1.e below.

6 Discussion

6.1 Summary of clinical effectiveness issues

The issues of clinical effectiveness raised in the original ERG report are unaffected by the further appraisal of the MS in the light of the responses to the second set of clarifications requested which related almost entirely to the validity of the economic model. These key issues can be summarised as:

- Strengths. The evidence was based on an adequately powered RCT, study AZA-001 reported by Fenaux et al in Lancet Oncology in 2009. The trial was registered and a protocol for the trial was available. The effects on survival, time to progression to AML, independence from transfusion and reduction in infections requiring intravenous antibiotics were clinically important and unlikely to have been explained by chance alone. There was no evidence that these benefits were substantially off-set by adverse events.

- Weaknesses. The AZA-001 study was open to bias, particularly from lack of blinding and uncertainty about losses to follow-up. In addition there was no direct evidence on impact on HRQoL. There is no evidence for differences in effects between investigator pre-selected treatment groups.

This addendum does also further consider in section 2 whether more bias has been added by the way in which the evidence arising from the AZA-001 study has been translated into input parameters in the MS economic model. As indicated there are considerable concerns that the methods used, particularly the choice of curve of best fit have introduced bias. Further the means by which information on time to progression to AML from the AZA-001 study was incorporated into the model is also highly problematic.

Finally in the original ERG report (section 5.4) we raised the issue that the face validity of the economic model was challenged by the mismatch between the observed improvement in median overall survival of 9.5 months in the AZA-001 trial

and the predicted improvement of 32-34 months in the MS model. We have further developed this argument, presented in Appendix 2, and suggest that the observation remains valid and strengthens the concerns raised in the detailed appraisal of the MS model presented in this addendum.

6.2 Summary of cost effectiveness issues

The model provided by the manufacturer and modified by the ERG is flawed and will be biased towards azacitidine on the basis that it underestimates mortality from AML. Given this, it overestimates survival, and thus is very likely to overestimate the potential benefits from treatment.

Within the ERG modified model, the treatment choice appears to favour BSC for each of the three groups considered in Study AZA-001. This finding appears reasonably robust to the different scenarios considered here but depends also on the ability of the model to compare outcomes across different types of treatments where there may be differences between groups.

The relationship between those preselected for SDC and LDC is relatively clear in the model. The SDC group typically contains those who are at elevated risk and who are younger, whilst the LDC group typically contains individuals with similar risks who we would expect to be older. Additionally, the manufacturer's submission suggested that SDC is considered for those who respond to remission-induction chemotherapy: this restriction does not necessarily apply to the LDC group. It would normally be expected these factors would bias findings *towards* SDC where both groups are compared against each other. Even with this possibly likely bias, LDC is preferred. Whether this suggests a lack of face validity of the model or raises serious questions as to current clinical practice may be an issue for the Committee to consider.

The relationship between the chemotherapy groups and the group receiving best supportive care is less clear. The BSC "group" appears to be not one group but two: those who are at less risk, and those who are at higher risks but who are not

candidates to receive chemotherapy. The analysis conducted here combines these groups because they were also combined within the submitted model based on Study AZA-001. It is not clear whether any bias between BSC and LDC/SDC would favour the former or the latter group. Without knowledge of this point, it is difficult to interpret the findings here.

Overall, the models before the committee – both the manufacturer's model and the ERG modifications to it – are likely to be subject to multiple biases of unknown effect. The lack of randomisation in Study AZA-001 reduces the practical usefulness of the study, and again it is the committee's role to judge the degree to which useful information can be gained from this study, and from any economic evaluation based upon it.

More broadly, one of the major issues with respect to this appraisal is the degree to which the ERG has been unable to verify critical elements of the model until late in the process, with the consequence that it has not been possible to validate some analyses conducted by the manufacturers. The degree of diagnosis and repair of the manufacturer's model has required that resources have been used in conducting tasks that would normally be the manufacturer's responsibility, at the expense of tasks that would normally form part of the ERG's role.

7 References

1. Kornblith, A. B. et al., 2002. Impact of Azacytidine on the Quality of Life of Patients With Myelodysplastic Syndrome Treated in a Randomized Phase III Trial: A Cancer and Leukemia Group B Study, *Journal of Clinical Oncology*, 20: 2441-2452.

2. McKenzie, L., and van der Pol, M., 2008. Mapping the EORTC QLQ C-30 onto the EQ-5D instrument: the potential to estimate QALYs without generic preference data. HERC, University of Aberdeen.

3. Sekeres, M. A. et al., 2004. Decision-making and quality of life in older adults with acute myeloid leukemia or advanced myelodysplastic syndrome, *Leukemia*, 18: 809-816.

4. Brazier J. E., Roberts J. R., 2004. The estimation of a preference-based index from the SF-12. *Medical Care*, 42: 851-859.

5. Franks, P., et al., 2004. Mapping the SF-12 to the EuroQol EQ-5D Index in a National US Sample. *Medical Decision Making* 24: 247-254.

6. Gray, A. M., Rivero-Arias, O., Clarke, P. M., 2006. Estimating the Association between SF-12 Responses and EQ-5D Utility Values by Response Mapping. *Medical Decision Making* 26: 18-29.

Appendix 1.

a. Overall survival plots for the modelled subgroups

The survival curves were derived using data presented in the manufacturer's model. The table below compares the median survival times from the ERG analysis with those shown in MS Table 6.7 (page 42) and those provided in the published account of the AZA-001 trial (Fenaux *et al*).

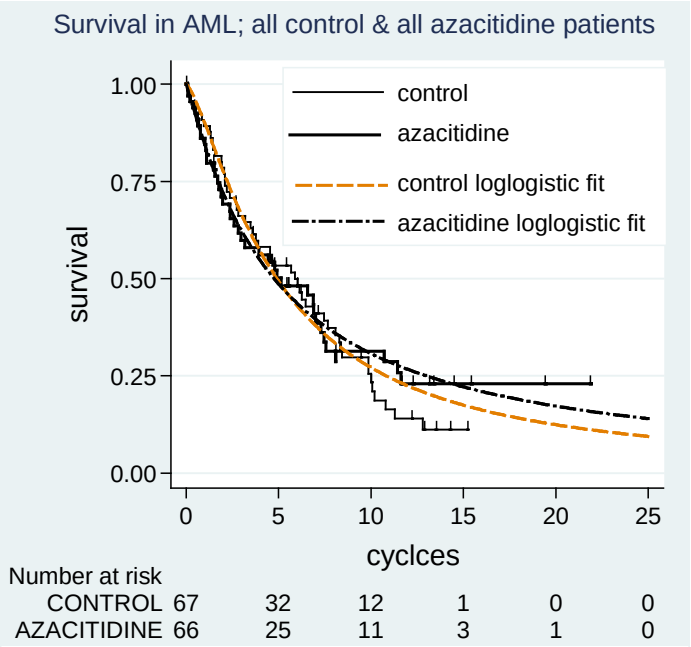
Submission	21.1	0.58	0.0045
Table 2 publication	21.1	0.58	0.0045
Data from model*	21.2	0.58	0.0025
Submission	11.5		
Table 2 publication	11.5		
Data from model*	11.6		
Submission	24.5	0.36	0.0006
Table 2 publication	24.5	0.36	0.0006
Data from model*	24.5	0.58	0.062
Submission	15.3		
Table 2 publication	15.3		
Data from model*	15.3		
Submission	25.1	0.76	0.51
Table 2 publication	25.1	0.76	0.51
Data from model*	25.2	0.62	0.24
Submission	15.7		
Table 2 publication	15.7		
Data from model*	15.8		

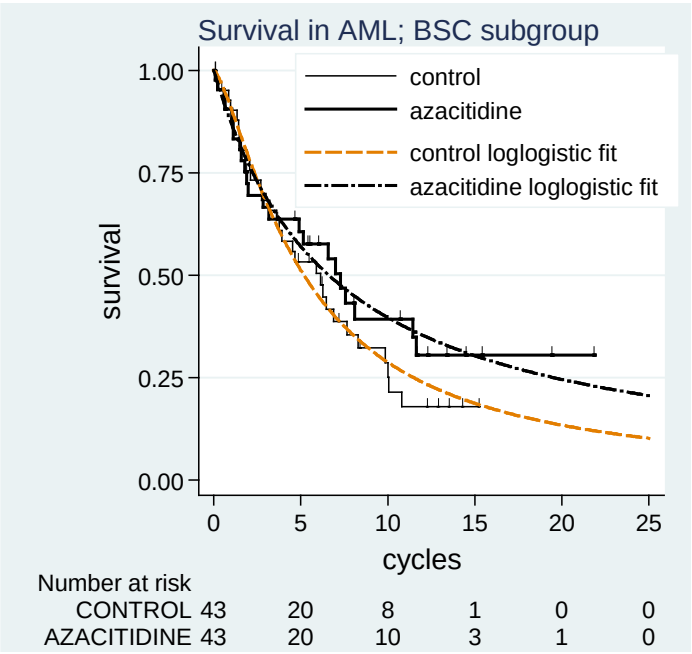
* Converted from cycles (defined as 5 weeks or 35 days) assuming 365/12 days per month.

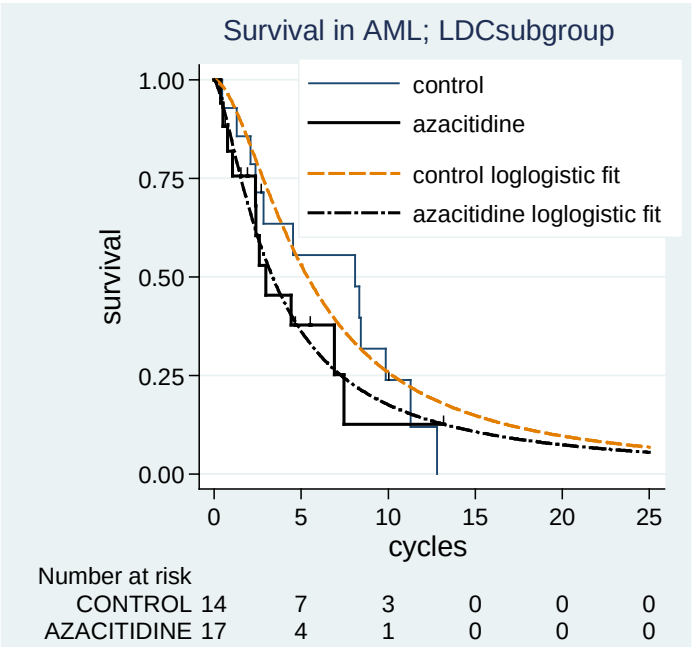
The small discrepancies in median times are probably due to inconsistencies in the data entries in the manufacturer's model. Hazard ratios in the publication and MS were calculated with a stratified analysis and this accounts for differences to HR derived from the model data. It was necessary to convert cycles into months in order to compare results between MM and MS; however months were not clearly defined and an assumption was made that 1 month in the MS and AZA-001 publication was equivalent to 365/12 days rather than 28 days.

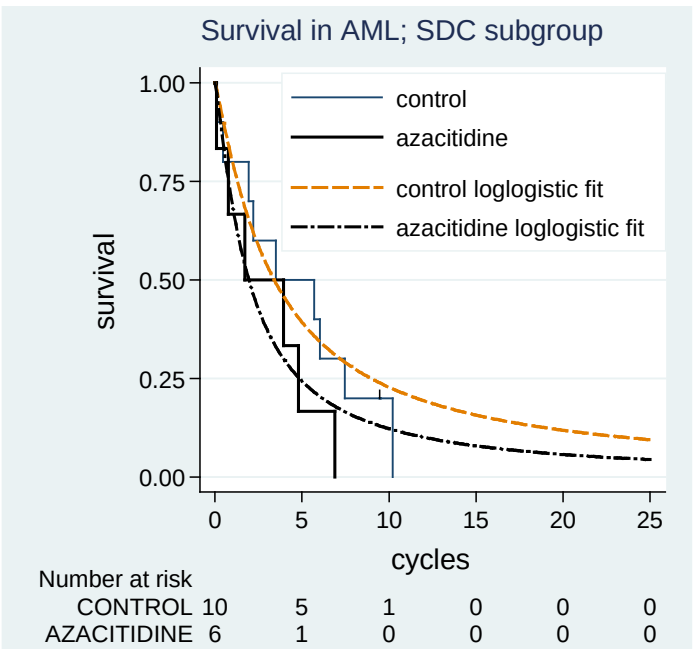
b. Observed survival in AML state

Data for survival in the AML state was extracted from the manufacturer's model and Kaplan-Meier plots prepared as shown below.


Survival in AML; all control & all azacitidine patients
survival
1.00
0.75
0.50
0.25
0.00
0
5
10
15
20
25
cyclces
control
azacitidine
control loglogistic fit
azacitidine loglogistic fit
Number at risk
CONTROL 67 32 12 1 0 0
AZACITIDINE 66 25 11 3 1 0


Survival in AML; BSC subgroup
survival
1.00
0.75
0.50
0.25
0.00
0
5
10
15
20
25
cycles
control
azacitidine
control loglogistic fit
azacitidine loglogistic fit
Number at risk
CONTROL 43 20 8 1 0 0
AZACITIDINE 43 20 10 3 1 0


Survival in AML; LDCsubgroup
survival
1.00
0.75
0.50
0.25
0.00
0
5
10
15
20
25
cycles
control
azacitidine
control loglogistic fit
azacitidine loglogistic fit
Number at risk
CONTROL 14 7 3 0 0 0
AZACITIDINE 17 4 1 0 0 0


Survival in AML; SDC subgroup
survival
1.00
0.75
0.50
0.25
0.00
0
5
10
15
20
25
cycles
control
azacitidine
control loglogistic fit
azacitidine loglogistic fit
Number at risk
CONTROL 10 5 1 0 0 0
AZACITIDINE 6 1 0 0 0 0

c. Modelled fits to observed survival in AML

The model base case assumed survival in AML can be plausibly described by a log-logistic distribution (shape parameter as for overall survival and scale parameter adjusted from that for overall survival). A potential test of this assumption is to compare various distributions fitted to the observed survival in AML using AIC criteria as described for overall survival on page 76 of the MS. The MS did not perform this comparison. The results below summarise AIC values for log-logistic, lognormal, Weibull, Gompertz and exponential models fit to the observed survival in AML for each of subgroup (best distribution shaded). For all subgroups the exponential or Gompertz distribution provided the superior AIC value and in no subgroup was the log-logistic distribution superior.

	Patient subgroup					
	SDC AZA	SDC CCR	LDC AZA	LDC CCR	BSC AZA	BSC CCR
Distribution						
Loglogistic	29.71486	43.94905	51.58938	45.59319	123.5243	126.9621
Lognormal	29.65329	44.57002	51.20741	45.44641	123.3296	128.6689
Weibull	27.84117	41.74195	52.30397	42.89394	124.6729	126.3899
Gompertz	27.25417	41.38423	52.16522	41.35668	123.3717	126.5135
Exponential	25.84184	39.79177	50.31894	42.54913	124.5116	124.5234

d. Time to progression to AML

The unpublished AZA-001 trial report states that bone marrow assessment was [redacted] The table below is taken from the trial report and shows the actual number and the potential scheduled number of assessments in each arm of the trial.

Table 11-15

[redacted]

The [redacted] render the estimate of TTP problematical and at risk of bias.

[redacted] the results of which are summarised below.

Type of analysis (data from AZ-001 trial report); median times are months

	[redacted]	[redacted]	[redacted]
AZACITIDINE	[redacted]	[redacted]	[redacted]
CCR	[redacted]	[redacted]	[redacted]
Difference	[redacted]	[redacted]	[redacted]

These analyses cannot remedy the problem of [redacted]. In summary there is risk of bias and considerable uncertainty associated with the estimate of TTP and any comparison between treatments and this is likely to equally affect an estimate of survival in AML.

e. Additional ERG economic analysis

Analysis of ERG model using Weibull survival analysis

Where Weibull analyses are used, the results suggest a lower cost-effectiveness of azacitidine when used as an adjunct to existing therapies. Here, azacitidine provides benefits at £67,243 per QALY when compared to BSC, at £64,792 per QALY when compared to LDC and at £44,627 when compared to SDC.

Table 3 Summary of cost-effectiveness results: Weibull survival curves (deterministic)

Treatment option	Outcomes		Marginal Differences		ICER (adjunct therapy)
	Costs	QALYs	Costs	QALYs	
Azacitidine BSC	£96,637	2.01	£66,954	1.00	**£67,243**
BSC	£29,683	1.01			
Azacitidine LDC	£98,251	2.03	£62,092	0.96	**£64,792**
LDC	£36,159	1.08			
Azacitidine SDC	£89,147	1.64	£30,307	0.68	**£44,627**
SDC	£58,840	0.96			

Key: BSC: best supportive care; LDC: low-dose chemotherapy; SDC: standard-dose chemotherapy.

Of all the analyses, the Weibull analysis is the only one in which an SDC option is cost-effective across the £0-100k per QALY range. For values of the threshold below £54,174, BSC provides the greatest net benefit. For values above this threshold, the azacitidine SDC option provides the greatest benefit.

Within this analysis, the second most cost-effective outcome in range where azacitidine SDC is provides the most net benefit is azacitidine LDC. As such, it is unsurprising that this provides the most net benefit in the comparison of BSC and LDC options. For this case, BSC is cost-effective up to £54,174 per QALY, with azacitidine LDC providing the most net benefit beyond this point.

The cost-effectiveness frontiers for these two cases are given in Figure 5 and Figure 6 below.

Figure 5. Cost-Effectiveness Frontier: SDC Options

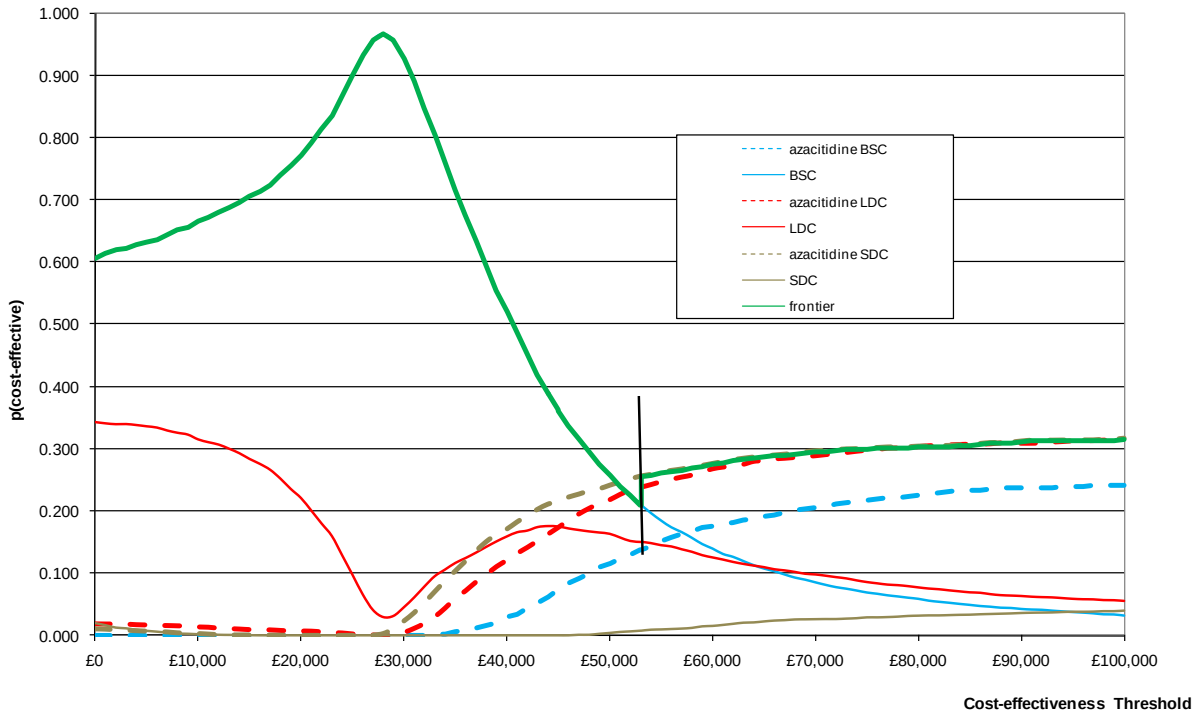
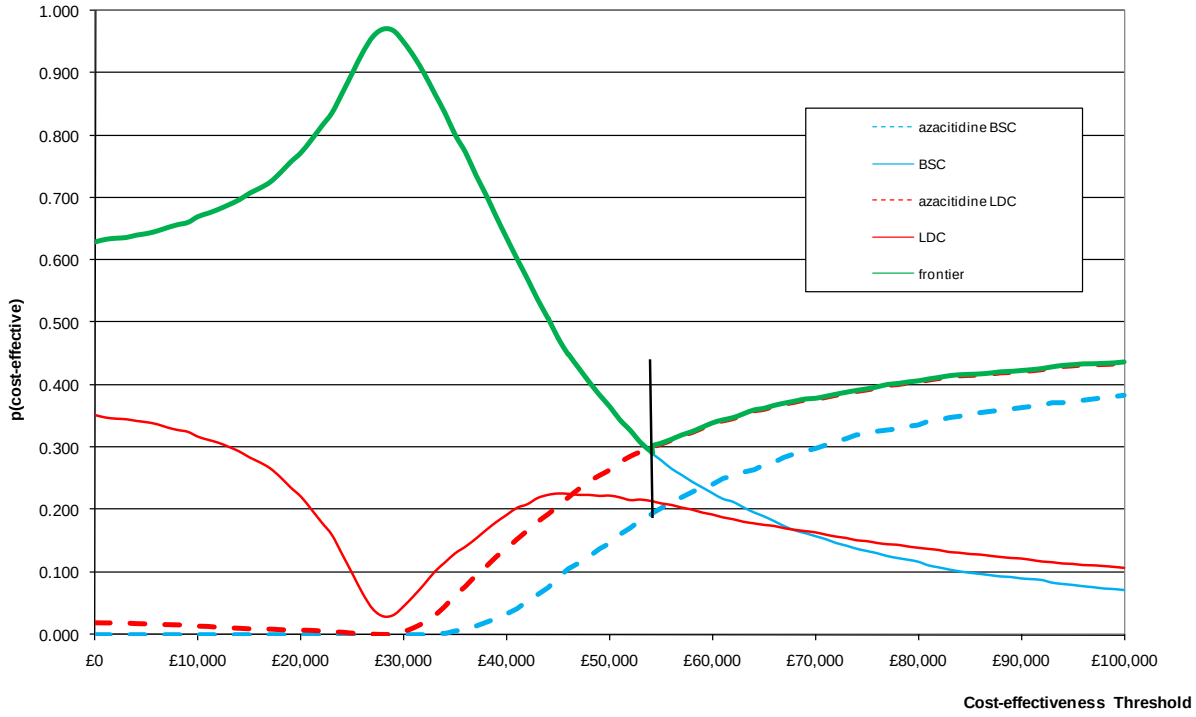


Figure 6. Cost-Effectiveness Frontier: LDC Options



For the BSC versus BSC plus azacitidine case, adding the treatment is not cost-effective in the general range of £20,000-£30,000 per QALY but may be cost-

effective at higher values. At the deterministic ICER of £67,423 per QALY, 50% of samples recommend adding azacitidine to BSC.

Figure 7. Cost-Effectiveness Frontier: BSC + azacitidine (vs BSC)

Analysis of ERG model using age/gender-specific mortalities

Where age and gender specific mortality figures are incorporated into the model, the analysis suggests a analyses are used, the results suggest a lower cost-effectiveness of azacitidine when used as an adjunct to existing therapies. Here, azacitidine provides benefits at £67,243 per QALY when compared to BSC, at £64,792 per QALY when compared to LDC and at £44,627 when compared to SDC.

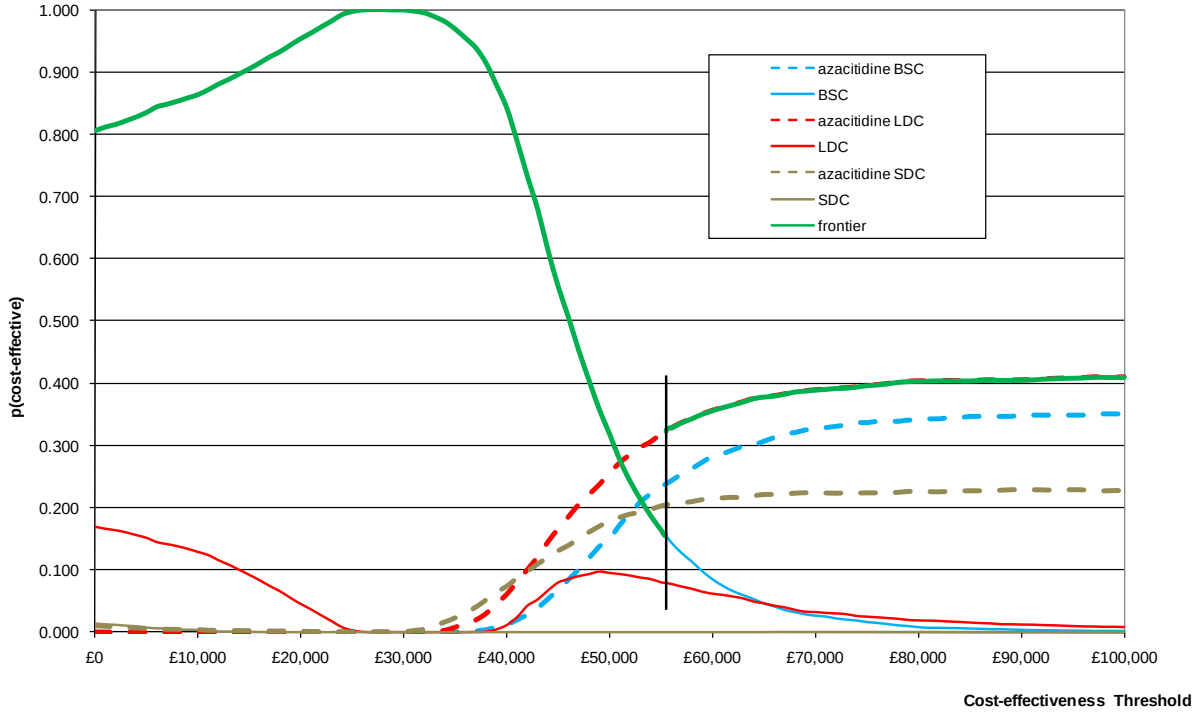
Table 4. Summary of cost-effectiveness results: mortality (deterministic)

Treatment option	Outcomes		Marginal Differences		ICER (adjunct therapy)
	Costs	QALYs	Costs	QALYs	
Azacitidine BSC	£112,923	2.65	£76,036	1.37	**£55,644**
BSC	£36,887	1.28			
Azacitidine LDC	£115,497	2.72	£66,516	1.26	**£52,613**
LDC	£48,981	1.45			
Azacitidine SDC	£104,844	2.26	£41,133	1.15	**£35,723**
SDC	£63,710	1.11			

Key: BSC: best supportive care; LDC: low-dose chemotherapy; SDC: standard-dose chemotherapy.

The cost-effectiveness frontier suggests similar findings to the log-logistic model, with BSC optimal over an additional £4000 per QALY. Here, BSC is optimal for all options up to £55,407, with azacitidine and LDC optimal thereafter.

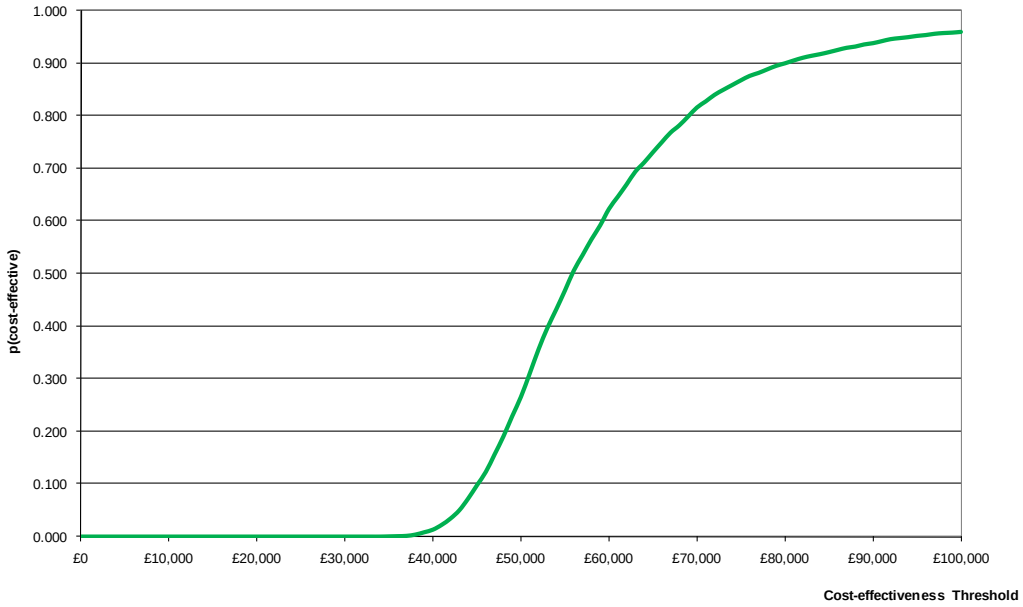
Figure 8. Cost-Effectiveness Frontier: SDC Options



As the SDC options are not optimal, the recommendations for the preselected LDC group are as for the preselected SDC group. In the BSC group the most cost-effective option is again BSC. At £55,644 per QALY, BSC plus azacitidine is cost-effective in 49% of cases sampled.

Overall, the incorporation of age-related mortality in the later periods of the model suggests that the both the costs and benefits of treatments are overstated. Overall, the impact of higher mortality appears to be to worsen the cost-effectiveness of azacitidine based treatments (as an adjunct to standard care) and increase the range over which BSC would be cost-effective in wider comparisons of treatment options.

Figure 9. Cost-Effectiveness Frontier: BSC + azacitidine (vs BSC)



The model as provided (and modified) underestimates AML mortality and typically assumes no MDS mortality (between 94-96% of cycles in the ERG baseline model had no MDS mortality and an underestimated AML mortality). Within this analysis, the addition of age-specific mortality appears to worsen cost-effectiveness, and it is feasible that corrected MDS and AML mortality figures may do the same.

Analysis of ERG model without correlations

The deterministic results for this scenario are as for the baseline scenario, since the assumption only applies to the sensitivity analysis. As such, we expect that the ICER for adding azacitidine to the three original therapies would be £51,973 (BSC), £48,766 (LDC) and £34,525 (SDC) per QALY.

For the probabilistic sensitivity analysis on the SDC group (Figure 10) the results are similar to baseline, with BSC optimal up to £52,644 per QALY and azacitidine LDC optimal thereafter. Again, since neither SDC option is cost-effective, the LDC options provide the same recommendations. For the BSC options (Figure 11), the CEAC is almost identical to the original model.

Figure 10. Cost-Effectiveness Frontier: SDC Options

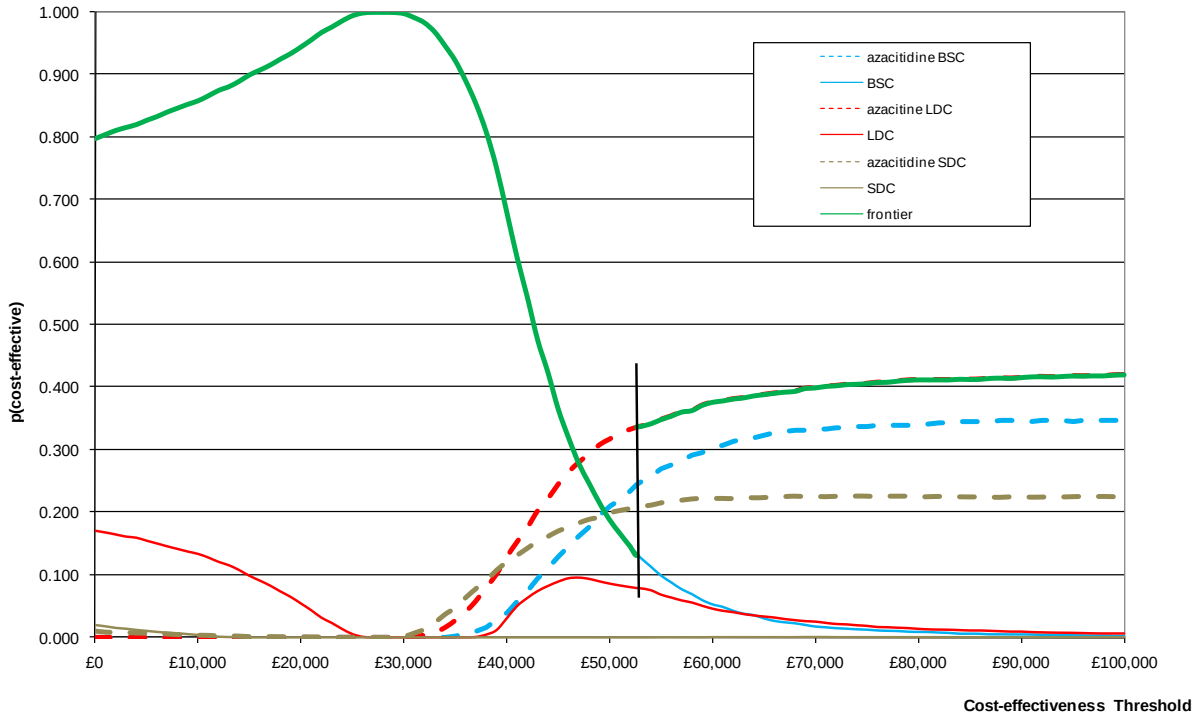
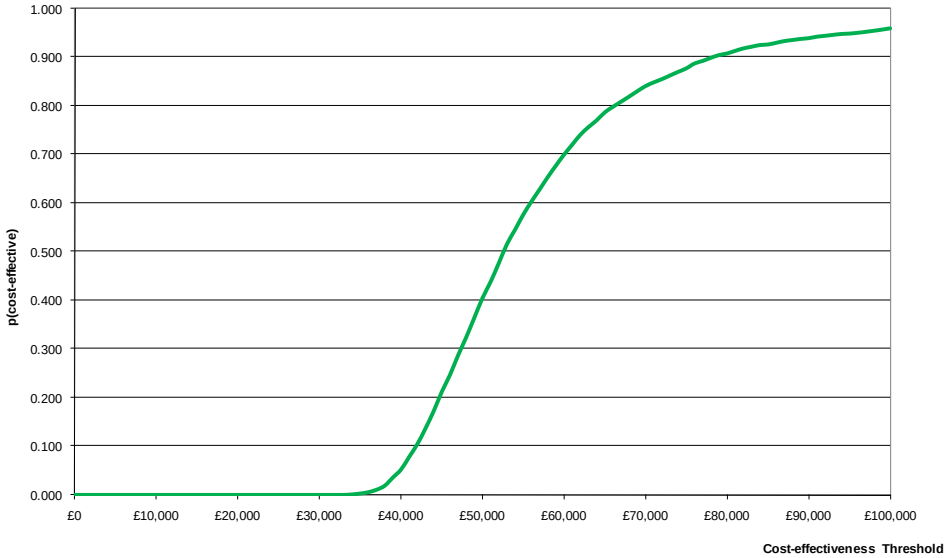


Figure 11. Cost-Effectiveness Frontier: BSC + azacitidine (vs BSC)



Appendix 2. Further argument concerning face validity of results of MS model relative to clinical effectiveness evidence

In the first ERG report we commented on issues of face validity in section 5.4:

> The validity of the results is severely undermined by the problems with the model indicated above. These concerns are reinforced by issues with the face validity of the results of the model relative to the results of the main source of evidence on clinical effectiveness. Thus the trial by Fenaux et al with median follow-up of 21.1 months indicates:
>
> - An improvement in median survival of 9.5 months with azacitidine
> - Little evidence of greater chance of cure with azacitidine
> - Likely very low levels of long term survival irrespective of treatment
>
> In contrast the model presented in the MS suggests an improvement in mean survival of 32 to 34 months (see MS table 7.18), considerably different from the observed difference, notwithstanding that one is a median and the other a mean. In a life-time model, although this difference might be explicable if azacitidine was bringing about an improvement in cure rate, such improvement in cure is not compatible with the trial results, nor indeed is it claimed. Thus the difference between modelled and observed survival time, particularly its magnitude, deserves explanation and seems likely on the basis of the observed errors in the model to be mainly because the model is not performing in the way envisaged.

The following diagrams further illustrate the nature of the issue raised and particularly address whether difference between the survival times observed in the trial and those predicted in the MS are merely explained by the fact that the results in the trial are reported as medians and those in the model as means.

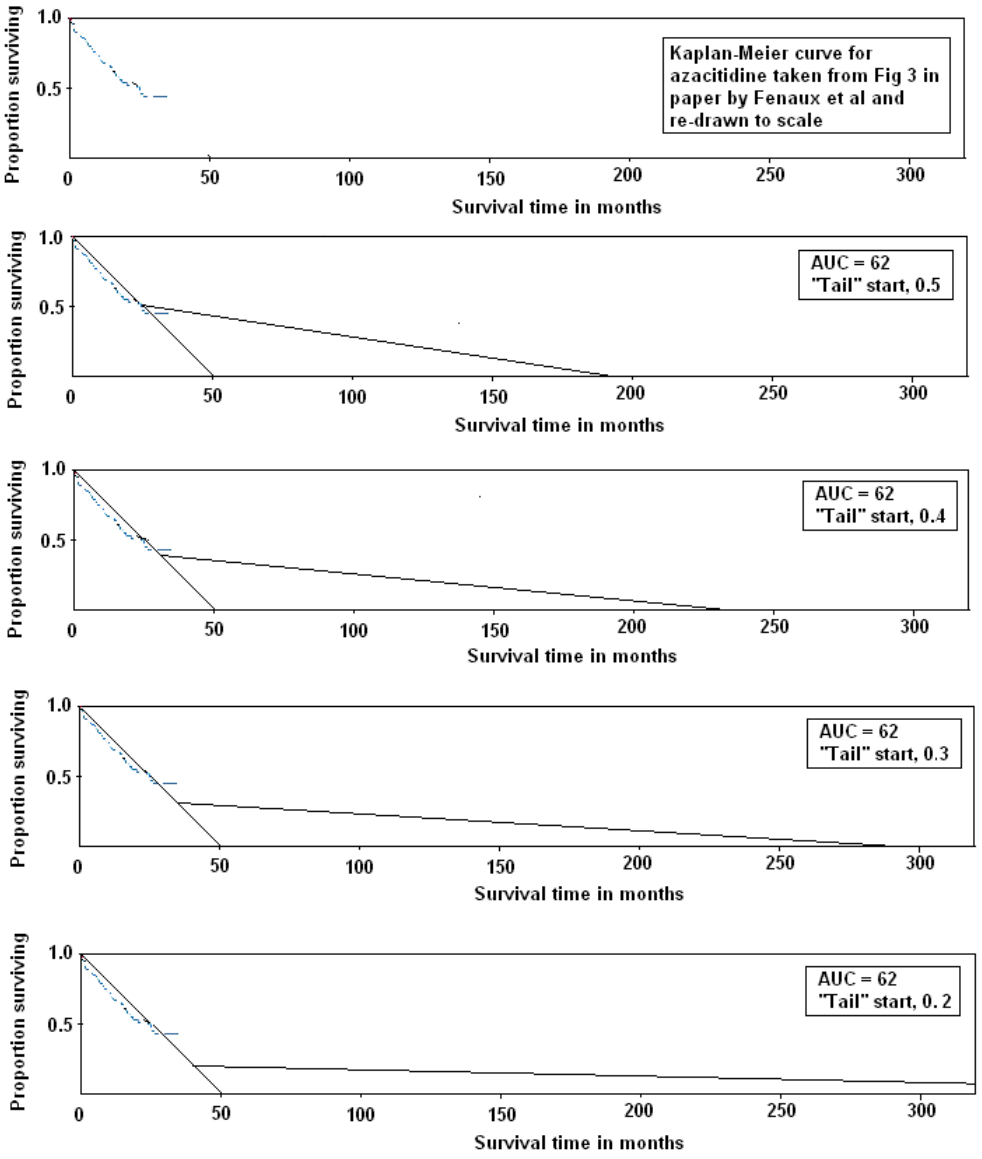
All the diagrams relate to the overall survival times for azacitadine:

- In the RCT by Fenaux et al, the median reported survival time is 24.5 months. The underlying Kaplan-Meier curve is also produced as Figure 3 in the Lancet paper
- In the original MS model the mean reported overall survival time as reported in Table 7.18 was 61.59 months. This value remains unchanged in the revised base-case offered in the response to the second set of issues for clarification

The diagrams below superimpose the actually observed survival curve for azacitidine on a number of illustrative curves compatible with a mean modelled survival time of 62 months. These illustrative curves are generated on the following bases:

- The area under a survival curve is equivalent to the mean survival
- The area under each survival = 62 calculated on geometrical bases
- The initial portion of each of the illustrative curves is based on an approximation to a triangle of the actual survival curve – this is the same for each illustrative curve
- The second portion of each curve considers that the “tail” of the curve applies to different proportions of the population: 0.5; 0.4; 0.3; and 0.2.

Figure A2. Illustrative survival curves compatible with the MS modelled mean overall survival time of 62 months



Comparing the shape of the illustrative curves with the actual observed results of the trial by Fenaux et al supports the view that the results of the model lack face validity. In order to generate an AUC of 62, the implied length of “tail” in each illustrative case suggests additional survival which is barely compatible with normal life expectancy in the general population, let alone in a population with high risk MDS. The approximate mean age at enrolment of patients in the trial by Fenaux et al was 70y.

It thus seems that the model retains patients far longer than would be expected. This not only affects benefits, but also costs. So it may explain another feature of the MS model results which is difficult to understand which is the magnitude of the additional costs associated with azacitidine treatment (£122k vs £38k [aza vs BSC, as reported in Table 7.18 in the response to the second set of issues for clarification]) far exceeding the additional costs attributable to acquisition of the drug which are in the region of £40-50k. Because both benefit and cost is affected by retention of patients in the model it is difficult to predict from this particular consideration whether the overall cost-effectiveness has currently been over- or under-estimated.