Blinatumomab for acute lymphoblastic leukaemia for people with minimal residual disease activity in remission: A Single Technology Appraisal

Produced by	School of Health and Related Research (ScHARR), The University of Sheffield
Authors	Jean Hamilton, Research Fellow, ScHARR, University of Sheffield, UK Daniel Pollard, Research Associate, ScHARR, University of Sheffield, UK Emma Simpson, Senior Research Fellow, ScHARR, University of Sheffield, UK Paul Tappenden, Reader in Health Economic Modelling, ScHARR, University of Sheffield, Sheffield, UK Ruth Wong, Information Specialist, ScHARR, University of Sheffield, Sheffield, UK Joanna Leaviss, Research Fellow, ScHARR, University of Sheffield, UK Clare Rowntree, Consultant Haematologist, University Hospital of Wales, Cardiff, UK Tobias Menne, Consultant Haematologist, Freeman Hospital, Newcastle, UK
Correspondence Author	Jean Hamilton, ScHARR, University of Sheffield, UK

Date completed	30th January 2018

Source of funding: This report was commissioned by the NIHR HTA Programme as project number 16/112/08.

Declared competing interests of the authors

Tobias Menne is a member of three independent data monitoring committees for studies sponsored by Amgen, but for different interventions than that considered in the current appraisal. None of the other authors have any conflicts of interest to declare.

Acknowledgements

We would like to thank Sarah Davis, Senior Lecturer, ScHARR, for providing comments on the draft report and Gill Rooney, Programme Manager, ScHARR, for providing administrative support and in preparing and formatting the report.

Rider on responsibility for report

The views expressed in this report are those of the authors and not necessarily those of the NIHR HTA Programme. Any errors are the responsibility of the authors.

This report should be referenced as follows:

Hamilton J, Pollard D, Simpson EL, Tappenden P, Wong R, Leaviss J, Rowntree C, Menne T. Blinatumomab for acute lymphoblastic leukaemia for people with minimal residual disease activity in remission: A Single Technology Appraisal. Final report to the National Institute for Health and Care Excellence. School of Health and Related Research (ScHARR), University of Sheffield, 2018.

Contributions of authors

Emma Simpson and Joanna Leaviss summarised and critiqued the clinical effectiveness evidence reported within the company's submission. Jean Hamilton critiqued the statistical analyses undertaken by the company. Paul Tappenden and Daniel Pollard critiqued the health economic analysis submitted by the company. Ruth Wong critiqued the company's search strategy. Clare Rowntree and Tobias Menne provided clinical advice to the ERG throughout the project. All authors were involved in drafting and commenting on the final report.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF BOXES

Abbreviations

AE	Adverse event
ANC	Absolute neutrophil count
AIC	Akaike information criterion
ALL	Acute lymphoblastic leukaemia
AML	Acute myeloid leukaemia
ATE	Average treatment effect
ATT	Average treatment effect on the treated
BCP	B-cell precursor
BIC	Bayesian Information Criterion
BNF	British National Formulary
BSA	Body surface area
CEAC	Cost-effectiveness acceptability curves
CI	Confidence interval
CMU	Commercial Medicines Unit
CMV	Cytomegalovirus
CNS	Central nervous system
CR	Complete remission
CRD	Centre for Reviews and Dissemination
CS	Company's submission
CSF	Cerebrospinal fluid
DCAS	Direct comparison analysis set
DES	Discrete event simulation
DFS	Disease-free survival
DSA	Deterministic sensitivity analyses
ECOG	Eastern Cooperative Oncology Group
EFS	Event-free survival
EMBASE	Excerpta Medica dataBASE
eMIT	Electronic Market Information Tool
EORTC QLQ-C30	European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire C30
EPAR	European Public Assessment Report
EQ-5D	Euroqol 5-Dimensions
ERG	Evidence Review Group
FCE	Finished consultant episode
GEE	Generalised estimating equation
GLM	Generalised linear model
GMALL	German Multicenter Acute Lymphoblastic Leukaemia Study Group
HB	Haemoglobin
HR	Hazard ratio
HRQoL	Health-related quality of life

HSCT	Haematopoietic stem cell transplantation
HST	Highly Specialised Technologies
ICER	Incremental cost-effectiveness ratio
IPD	Individual patient-level data
IPTW	Inverse probability of treatment weighting
IV	Intravenous
Kg	Kilogram
MEDLINE	Medical Literature Analysis and Retrieval System Online
Mg	Milligram
MRD	Minimal residual disease
NHS	National Health Service
NICE	National Institute for Health and Care Excellence
NIH	National Institutes of Health
NSAIDS	Non-steroidal anti-inflammatory drugs
OS	Overall survival
PAS	Primary analysis set
PCR	Polymerase chain reaction
Ph	Philadelphia chromosome
PRO	Patient-reported outcome
PRS	Post-relapse survival
PSA	Probabilistic sensitivity analyses
PSS	Personal Social Services
QALY	Quality-adjusted life year
QuEENS	Quality of Effectiveness Estimates from Non-randomised Studies
RCS	Restricted cubic spline
RCT	Randomised controlled trial
RFS	Relapse-free survival
ROBINS	Risk Of Bias In Non-randomised Studies
R/R	Relapsed/refractory
RT-qPCR	Real-time quantitative polymerase chain reaction
SAE	Serious adverse event
sATT	Stabilised average treatment effect on the treated
SC	Standard care
SD	Standard deviation
SE	Standard error
TA	Technology appraisal
TCR	T-cell receptor
TKI	Tyrosine kinase inhibitor
WBC	White blood cells
WTP	Willingness-to-pay

1. SUMMARY

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

The company's submission (CS) assesses the clinical effectiveness and cost-effectiveness of blinatumomab (Blincyto®), within its anticipated licensed indication for the treatment of adult patients with minimal residual disease-positive B-cell precursor acute lymphoblastic leukaemia (MRD+ BCP-ALL) whilst in remission. The company's description of ALL and its management is broadly appropriate. The decision problem addressed by the CS is partly in line with the final scope issued by the National Institute for Health and Care Excellence (NICE). The indirect comparison and health economic analysis presented within the CS compare blinatumomab with standard care chemotherapy within a population of adult patients with Philadelphia chromosome-negative (Ph-) disease with first complete haematological remission (CR1); this is narrower than the population defined by the anticipated license indication for blinatumomab. As such, the company's indirect comparison and health economic analysis exclude two groups of patients who were enrolled into the BLAST study: (i) patients who are in second or subsequent haematological remission (CR2+), and (ii) patients with Ph+ ALL (any CR). Despite this absence of evidence, the CS argues that due to the substantial unmet need across all subgroups, blinatumomab should be considered for use within its full anticipated marketing authorisation. However, the company further suggests that blinatumomab should be used early in the treatment pathway, with initiation after front-line chemotherapy (after two induction cycles) for those patients with persistent MRD at this stage. The CS also excludes the comparator of "monitor for relapse" based on the argument that it is highly unlikely that MRD+ patients who are at high risk of relapse would not receive active treatment. However, clinical advisors to the Evidence Review Group (ERG) noted that due to its favourable toxicity profile, blinatumomab may be a potential treatment option for patients who are unable to undergo haematopoietic stem cell transplantation (HSCT) or to tolerate chemotherapy; the ERG considers that a further comparison of blinatumomab versus monitoring within this subgroup should have been explored.

1.2 Summary of clinical effectiveness evidence submitted by the company

The key clinical effectiveness evidence for blinatumomab was based on two single-arm open-label studies; BLAST (n=116) and the pilot study MT103-202 (n=20). From the 116 patients in BLAST, median overall survival (OS) was [redacted], with an 18-month OS probability of [redacted]. From 110 patients providing relapse-free survival (RFS) data from BLAST, median RFS was [redacted], with an 18-month RFS probability of [redacted]. Based on the European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire C30 (EORTC QLQ-C30), patients reported [redacted]. By the end of the core study,

[redacted]

HRQoL as measured by the Euroqol 5-Dimensions (EQ-5D) questionnaire did not change significantly from baseline to the end of the core study. [redacted] experienced at least one treatment-emergent AE.

Comparator data relating to standard care chemotherapy were provided from one historical control study, Study 20120148 (n=287); this study was based on data obtained from existing clinical databases.

Owing to the lack of randomised data to inform the comparative effectiveness of blinatumomab versus standard care chemotherapy, treatment effects were estimated using non-randomised data from BLAST and the historical control study. Due to differences between the populations of BLAST and the historical control study, comparative analyses were undertaken using subsets of the original study populations which were restricted to patients with Ph- disease in CR1 only: the BLAST primary analysis set (PAS, [redacted]) and the historical control direct comparison analysis set (DCAS, [redacted]). A propensity score model was constructed and used to generate weights which were applied to the historical control DCAS, with the aim of approximating the response to standard care chemotherapy that would be expected in a population with the same characteristics as the BLAST PAS. The resulting average treatment effect on the treated (ATT) estimates are applicable to Ph- and CR1 individuals only. This analysis suggested a hazard ratio (HR) [redacted].

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

Despite limitations in the company's search strategy, the ERG considers it unlikely that any relevant studies of blinatumomab in adult BCP-ALL patients with MRD positivity after treatment have been missed by the company's searches. The eligibility criteria applied in the selection of evidence for the clinical effectiveness review were considered by the ERG to be reasonable and consistent with the decision problem outlined in the final NICE scope, with the exception that the comparator "monitor for relapse" was not included. The ERG's clinical advisors noted that some older and less fit patients may not be able to receive HSCT or to tolerate chemotherapy, but may be able to tolerate blinatumomab, and so this comparator is potentially relevant for a subgroup of MRD+ BCP-ALL patients. It is unclear whether potentially relevant comparator data exist for this subgroup (for example, from registry sources).

The main evidence in the CS was from the single-arm BLAST study. Whilst BLAST was generally well reported and conducted, single-arm studies are associated with an array of potential biases including a high risk of selection bias (due to the absence of randomisation), performance bias and detection bias (due to the absence of blinding).

The ERG considers that the propensity score methods used by the company to inform comparative effectiveness estimates were appropriate. However, the estimation of treatment effects based on non-randomised data is still subject to inherent limitations, namely that it is not possible to account for unobserved confounders. It was unclear whether the uncertainty associated with the propensity score weights was accounted for when estimating the treatment effects. The ERG therefore considers that the reported treatment effects are likely to underestimate the associated uncertainty and should be interpreted with caution. There was also a lack of clarity and consistency in the weighted analyses presented within the CS, as results using stabilised ATT (sATT) weights were presented in the clinical effectiveness section, and standard (non-stabilised) weights were used to inform the health economic model.

The key uncertainties in the clinical evidence relate to the comparative efficacy and the generalisability of the available evidence to the full population outlined in the scope.

1.4 Summary of cost effectiveness submitted evidence by the company

The company's *de novo* partitioned survival model assesses the cost-effectiveness of blinatumomab versus chemotherapy (based on the UK ALL14 maintenance regimen) in patients with Ph- MRD+ BCP-ALL in CR1. Incremental health gains, costs and cost-effectiveness of blinatumomab are evaluated over a 50-year time horizon from the perspective of the National Health Service (NHS) and Personal Social Services (PSS). The company's model is comprised of a main structure which reflects RFS and OS outcomes, together with two linked sub-models which are intended to estimate additional costs and HRQoL decrements associated with HSCT received before and/or after relapse. The main model structure includes three health states: (1) relapse-free; (2) post-relapse and (3) dead. The survival models were generated from analyses of time-to-event data (RFS and OS) from the company's propensity score analysis of the BLAST PAS and the historical control study DCAS using ATT weights. RFS is modelled using an unrestricted Gompertz distribution (equivalent to fitting separate models to both groups), whilst OS is modelled using a log normal mixture cure model (whereby the parameters of the log normal distribution are the same for both groups, but the cure fraction is allowed to differ between the groups). HRQoL is assumed to be principally determined by relapse status, time spent in the relapse-free state and treatment received; utility estimates were derived from a generalised linear model/generalised estimating equation (GLM/GEE) model fitted to EQ-5D data collected in BLAST, a further propensity matching analysis of the BLAST and TOWER blinatumomab studies, as well as other literature and assumptions. Resource use estimates and costs were based on data collected in BLAST, the UK ALL14 treatment protocol, routine cost sources, clinical opinion and other literature.

Based on the probabilistic version of the model (assuming the unrestricted Gompertz function for RFS and the log normal mixture cure model for OS), blinatumomab is expected to generate an additional 2.85 quality-adjusted life years (QALYs) at an additional cost of £84,456 compared with standard care: the corresponding incremental cost-effectiveness ratio (ICER) for blinatumomab versus standard care is £29,673 per QALY gained. The deterministic version of the company's model produces a similar ICER of £28,524 per QALY gained for blinatumomab versus standard care. Assuming a willingness-to-pay (WTP) threshold (λ) of £20,000 per QALY gained, the company's model suggests that the probability that blinatumomab produces more net benefit than standard care is 0.10; assuming a WTP threshold of £30,000 per QALY gained, the company's model suggests that the probability that blinatumomab produces more net benefit than standard care is 0.53. Following the clarification process, the company submitted a revised model which addressed some of the minor concerns initially raised by the ERG; the probabilistic version of the company's updated model suggests that the ICER for blinatumomab versus standard care is £28,655 per QALY gained.

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

The ERG critically appraised the company's economic analysis and double-programmed the deterministic version of the company's model. The ERG's critical appraisal identified several issues relating to the company's economic analysis and the evidence used to inform it. These include: (i) the exclusion of relevant patient subgroups from the model; (ii) the exclusion of the "monitor for relapse" comparator from the analysis; (iii) use of a model structure which is inappropriate for tracking HSCT; (iv) the absence of RCT evidence for blinatumomab versus standard care; (v) concerns regarding the company's approach to RFS/OS model selection; (vi) concerns regarding the robustness of the company's alternative base case (blinatumomab used on relapse for the standard care group); (vii) the questionable reliability of the company's HRQoL estimates; (viii) uncertainty surrounding the proportion of RFS events that are deaths; (ix) the inclusion of an unrealistic treatment pathway and (x) limited sensitivity analysis around alternative parametric functions.

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

1.6.1 Strengths

The ERG considers it unlikely that any relevant studies of blinatumomab have been missed the company's searches. The BLAST study was a well conducted single-arm study which reported on the full range of outcomes listed in the NICE scope (although comparative analyses from the company's propensity score model were restricted to RFS and OS outcomes only).

The CS details extensive efforts taken to verify the correct implementation of the health economic model and to ensure the accuracy of the parameter inputs against the source material from which these were derived. The company's model was found to include only minor errors.

1.6.2 Weaknesses and areas of uncertainty

The key weaknesses in the evidence base relate to the lack of randomised evidence to inform comparative effectiveness and the limited generalisability of the available evidence to the full population defined by the NICE scope and the anticipated license authorisation. The ERG considers the following to represent the key uncertainties within the clinical and economic evidence base for blinatumomab:

- The absence of comparative clinical and economic evidence for blinatumomab versus standard care chemotherapy within subgroups of the BLAST study which were excluded from the comparative analysis (patients with Ph+ MRD+ BCP-ALL and patients with Ph- MRD+ BCP-ALL with CR2+).
- The absence of clinical data and economic comparisons of blinatumomab versus monitoring for patients who are unable to undergo HSCT or to tolerate chemotherapy.
- The necessary reliance on adjusted historical control evidence, due to the absence of RCT evidence for blinatumomab versus standard care, and the potential for unobserved confounders.
- The long-term extrapolation of RFS and OS outcomes, including the timing of cure.

1.7 Summary of exploratory and sensitivity analyses undertaken by the ERG

The ERG undertook eight sets of exploratory analyses using the deterministic version of the company's updated model. Notwithstanding uncertainty relating to the choice of parametric RFS and OS functions, the ERG's preferred model includes the correction of seven minor programming errors and the inclusion of a fixed 5-year cure point. The ERG-preferred model produces a deterministic ICER for blinatumomab versus standard care of £30,227 per QALY gained. The ERG also undertook a number of further analyses to explore the impact of alternative parametric models and alternative parameter values on the results of the ERG-preferred model. These analyses indicate that the costs of standard care chemotherapy, the post-HSCT survival probabilities and the utility value for the post-relapse state have only a minor impact on the ICER for blinatumomab versus standard care. Conversely, the cure fraction and the choice of parametric OS distribution have a significant impact on the model results. Within the ERG's exploratory analysis of alternative RFS and OS models, the ICER for blinatumomab versus standard care ranges from £25,783 per QALY gained (Weibull non-mixture cure model, unrestricted) to £63,265 per QALY gained (Weibull model, unrestricted). Across the full range of models considered, only the Weibull non-mixture cure model (unrestricted) and the Weibull mixture cure model (unrestricted) produce results in which the full range of deterministic ICERs are below £30,000 per QALY gained (irrespective of RFS model assumed). The clinical advisors' three preferred

OS models (the generalised gamma [unrestricted], the restricted cubic spline (RCS) Weibull [unrestricted] and the Weibull mixture cure [unrestricted]) result in deterministic ICERs in the range £25,810 per QALY gained to £34,904 per QALY gained.

On the basis of the results of the 35 parametric OS models considered within the ERG's exploratory analyses, the ERG does not believe that blinatumomab meets NICE's criteria for life-extending treatments given at the end of life.

2. BACKGROUND

This report provides a review of the evidence submitted by Amgen in support of blinatumomab for acute lymphoblastic leukaemia (ALL) for people with minimal residual disease (MRD) activity in remission. It considers both the original company submission (CS)[1] received on 8th November 2017 and a subsequent response to clarification questions supplied by Amgen on 13th December 2017.

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

The CS[1] (pages 19-31) provides a reasonable description of the underlying health problem; this is summarised briefly below.

ALL is a rare and rapidly progressing form of leukaemia characterised by the excess production of immature lymphocyte precursor cells, called lymphoblasts or blasts cells, in the bone marrow. Lymphocytes are white blood cells that are vital for the body's immune system. Eventually, this affects the production of normal blood cells which leads to a reduction in the numbers of red cells, white cells and platelets in the blood.[1] ALL represents about 20% of all leukaemias in adults.[2, 3]

There are a number of sub-classifications of ALL, with the majority (approximately 76%) of adult cases being B-cell lineage (based on a weighted average of five estimates synthesised by the company[4, 5 6 7 8]). Of these, approximately 93% are B-cell precursor (BCP) ALL (based on a weighted average of two studies[9 10]). Therefore, BCP-ALL constitutes approximately 71% of the adult ALL population, which is expected to equate to around 236 patients in England and Wales.[1] Approximately 25%[3, 9] of adults with ALL (across all sub-classifications, not specifically BCP) have an acquired chromosomal abnormality, known as Philadelphia chromosome-positive (Ph+) disease, which is caused by reciprocal translocations between chromosomes 9 and 22. These translocations result in a BCR-ABL fusion gene that encodes an active tyrosine kinase protein which causes uncontrolled cell proliferation. The presence of the Ph chromosome in adults increases with age[2, 3, 11] and Ph+ ALL individuals typically have a worse prognosis than those without the abnormality.[12]

Many of the patients who achieve the criteria for haematological complete remission (CR) will experience a recurrence of disease; this is thought to result from residual leukaemia cells that remain.[1] MRD describes residual ALL in patients in CR that is detectable only by molecular means.[13] Patients are considered to have clinically significant MRD, and are described as being MRD+,[3, 14] if their MRD level is greater than 1×10^{-4}, although clinical studies have assessed MRD positivity using various thresholds. The company estimates that 36% of all BCP-ALL patients in CR exhibit MRD+, based on a weighted analysis of Ph- patients in three studies;[4, 15, 16] this implies an estimated 85 cases of MRD+

BCP-ALL in England and Wales.[1] The company's clarification response[17] (question A2) estimates that approximately 15 of these patients will be Ph+.

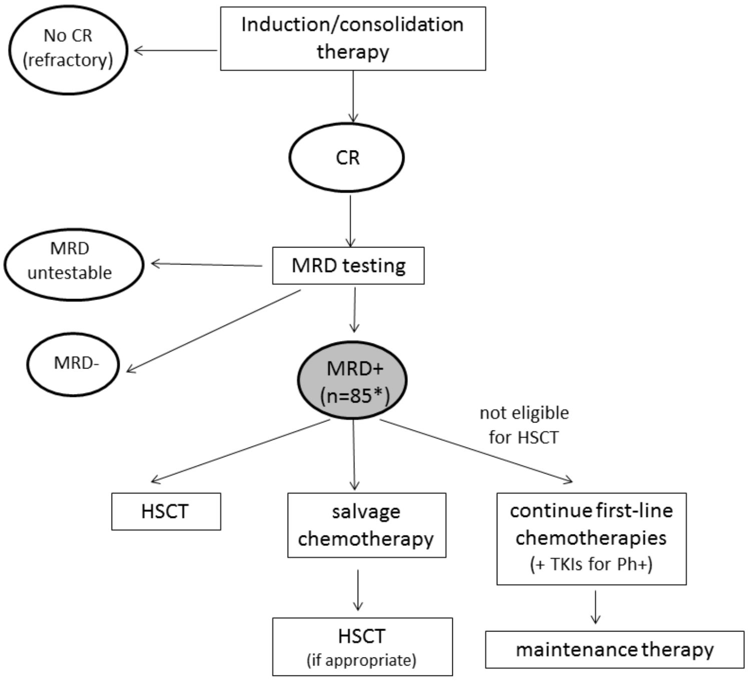
The prognosis for patients with BCP-ALL is dependent on a number of factors. Well-established positive prognostic factors include: younger age; shorter time to CR; longer duration of CR; absence of poor risk cytogenetics such as Ph+, and lower white blood cell counts.[3, 5, 7] In addition, MRD positivity is a major and well established risk factor.[13] In a large German Multicentre Acute Lymphoblastic Leukaemia (GMALL) study of Ph- ALL,[4] the probability of overall survival (OS) at 5 years was 42% for MRD+ compared with 80% for MRD- patients (MRD assessed at week 16 after consolidation therapy). In a meta-analysis by Berry *et al*,[18] poorer outcomes for MRD+ patients compared with MRD- patients were observed. Although OS estimates were not reported specifically for the MRD+ BCP-ALL subgroup, the persistence of MRD was shown to be a strong predictive factor for relapse and OS, irrespective of ALL cell phenotype (B-cell or T-cell), Ph chromosome subgroup and MRD detection method, cut-off, or timing of assessment.[18]

2.2 Critique of company's overview of current service provision

In general, the CS[1] (pages 26-30) provides a reasonable overview of current service provision for people with MRD+ BCP-ALL, although the submission is not always clear where information relates to specific sub-populations of ALL patients. The company's description of the treatment pathway is briefly summarised in this section, and is supplemented with information provided by clinical advisors to the Evidence Review Group (ERG).

The management of people with MRD+ BCP-ALL is complex and there is currently no guidance published by the National Institute for Health and Care Excellence (NICE) for the treatment of adults with MRD+ BCP-ALL in England. The treatment of ALL in the UK is generally based on the UKALL14 protocol.[19] In general, the treatment approach varies according to age, general fitness and health at diagnosis and the results of cytogenetic testing.[2] The aim of treatment is to achieve cure (defined as sustained MRD negativity) and maintained haematological CR (defined as a bone marrow blast level of <5%[13]). According to clinical advice received by the ERG, patients who do not experience relapse within 5 years of diagnosis are generally considered to be cured. Most long-term survivors achieve cure by undergoing haematopoietic stem cell transplantation (HSCT), although this is not always required for standard-risk patients, and some high-risk patients may not be suitable candidates for HSCT for various reasons, for example, due to older age, medical comorbidities, or the lack of a suitable donor.[13] Figure 1 presents an overview of the treatment pathway for people with MRD+ BCP-ALL.

Figure 1: **Overview of the treatment pathway for MRD+ BCP-ALL**



** Estimated number of patients in England and Wales from CS,[1] The grey box indicates the relevant population for this appraisal*

In clinical practice, most treatment plans have three phases: (i) induction (with or without intensification); (ii) consolidation, and (iii) maintenance (in adults, later stages of treatment may be replaced by allogeneic transplantation). The aim of induction therapy is to achieve full remission quickly. Patients are treated with established standard chemotherapy combinations (including tyrosine-kinase inhibitor [TKI] therapy for Ph+ patients only). Once in remission, patients may proceed to HSCT, with or without intensification, if considered high-risk for relapse (e.g. MRD+, poor risk cytogenetics, age over 40 years) assuming they are clinically eligible, willing to undergo HSCT and have a suitable donor. Currently, adult patients with a sibling donor would also undergo HSCT in first remission. Consolidation therapy followed by maintenance therapy is given to patients who are not eligible for HSCT or who have standard-risk disease and no sibling donor. Ph+ patients additionally receive daily imatinib (a TKI therapy) throughout induction and intensification. As patients with Ph+ disease are deemed high-risk, they would usually have an HSCT instead of ongoing consolidation and maintenance chemotherapy unless they are not considered fit enough for transplant or do not have a suitable donor.

MRD testing is widely implemented in the UK and is recommended as standard care in the patient management process for ALL.[13] However, global consensus has not yet been reached on when to test for MRD. Brüggemann *et al*[20] determined that the timing of MRD status influences outcomes, with patients achieving MRD negativity during induction experiencing improved relapse-free survival (RFS) and OS compared with patients achieving MRD negativity after induction. A survey of 20 UK physicians undertaken by the company[21] suggested an apparent consensus on MRD testing patterns in the UK. Based on the survey data, an initial prognostic MRD test was commonly conducted 4-8 weeks after the start of induction therapy. Once a patient has achieved MRD negativity, they do not have further testing if they remain on chemotherapy only. Patients undergoing transplantation will have an average of 4 post-CR MRD tests, at roughly 3 month intervals, over the subsequent 12 months post-transplant (irrespective of MRD status pre transplant). The rare patients that continue chemotherapy despite being MRD+ (due to patient choice, fitness or lack of donor) would not receive further routine MRD testing due to the lack of current options for curative treatment post-relapse.

The company suggests that blinatumomab should be used early in the treatment pathway, with initiation after front-line chemotherapy (after two induction cycles) for those patients with persistent MRD at this stage. According to the CS,[1] blinatumomab is expected to displace continued chemotherapy and/or be used prior to HSCT. Blinatumomab is not intended to displace HSCT, rather it is likely to be used prior to HSCT in patients who are eligible to undergo transplant, with the aim of increasing the likelihood of a positive outcome, or to delay the need for HSCT.[22] Despite this, the company suggests that by achieving and sustaining MRD negativity over time, blinatumomab may conceivably delay transplant indefinitely (the ERG notes that this argument suggests that blinatumomab would displace HSCT, at least in some patients).

3. CRITIQUE OF COMPANY'S DEFINITION OF THE DECISION PROBLEM

This chapter presents a summary and critique of the decision problem addressed by the CS.[1] A summary of the decision problem as outlined in the final NICE scope[22] and addressed in the CS[1] is presented in Table 1.

Table 1: Company's statement of the decision problem (reproduced from CS Table 1)

	Final scope issued by NICE[22]	Decision problem addressed in the CS[1]	Company's rationale if different from the final NICE scope
Population	People with BCP-ALL who have MRD activity while in haematological remission	Adults with MRD+ B-precursor ALL. Clinical evidence for blinatumomab is aligned with the proposed licensed indication; however, comparative evidence from a historical comparator study is limited to patients with Ph-negative B-precursor ALL who are in first complete haematological remission. Therefore, the economic analysis presented in this submission focused on this patient subgroup. Although the cost-effectiveness evidence does not consider the Ph+ population or later remission states, due to the substantial unmet need across all sub-populations blinatumomab should be considered for use in alignment with its full anticipated marketing authorisation.	Blinatumomab is not expected to have a marketing authorisation for use in paediatric patients in this indication.
Intervention	Blinatumomab	As per final scope	N/a
Comparator(s)	• Retreatment with combination chemotherapy • Monitor for relapse	• Retreatment with combination chemotherapy	Based on expert clinical opinion it is highly unlikely that MRD+ patients who have a high-risk of relapse would solely be monitored for relapse without any treatment. Therefore, in the economic evaluation monitoring for relapse is not considered a comparator in its own right – instead, it is captured alongside ongoing chemotherapy regimens.
Outcomes	The outcome measures to be considered include: • Overall survival • Disease-free survival • Relapse-free survival • MRD response • Rate of stem cell transplant • Adverse effects of treatment • HRQoL	As per final scope	N/a

	Final scope issued by NICE[22]	Decision problem addressed in the CS[1]	Company's rationale if different from the final NICE scope
Special considerations including issues related to equity or equality	If appropriate, the appraisal should include the costs associated with diagnostic testing for these cells in people with ALL, while in remission, who would not otherwise have been tested. A sensitivity analysis should be provided without the cost of the diagnostic test. Guidance will only be issued in accordance with the marketing authorisation. Where the wording of the therapeutic indication does not include specific treatment combinations, guidance will be issued only in the context of the evidence that has underpinned the marketing authorisation granted by the regulator.	As per final scope	MRD status testing is already routine clinical practice in the diagnostic work-up and monitoring of BCP-ALL,[13, 23] and is recognised as an important marker for informing treatment decisions and prognosis. No additional tests or investigations are required for treatment with blinatumomab.

N/a - Not applicable

3.1 Population

The population defined in the final NICE scope[22] relates to people with BCP-ALL who have MRD activity while in remission. Blinatumomab does not currently have a marketing authorisation for this indication. According to the draft SmPC submitted to NICE by the company,[24] the anticipated wording of the marketing authorisation is as follows: *"BLINCYTO [blinatumomab] is indicated for the treatment of adults with minimal residual disease (MRD) positive B precursor ALL."* This population is in line with the BLAST study,[1] but relates only to adult ALL patients. The ERG notes that the indirect comparison and the health economic analysis presented within the CS (see Sections 4.4 and 5.2, respectively) relate to a narrower population of adult patients with Ph- disease with first complete haematological remission (CR1). Consequently, the indirect comparison and health economic analysis exclude two groups of patients who were enrolled into BLAST and who are included in the anticipated marketing authorisation: (i) patients who are in second or subsequent haematological remission (CR2+), and (ii) patients with Ph+ ALL (any CR). Despite this absence of evidence, the CS argues that due to the substantial unmet need across all subgroups, blinatumomab should be considered for use within its full anticipated marketing authorisation. In addition, clinical advisors noted that due to its toxicity profile, blinatumomab represents a potential treatment option for patients who are unable to undergo HSCT or to tolerate chemotherapy; it is unlikely that this subgroup is reflected within the population of patients enrolled into the BLAST study. These issues are discussed further in Section 5.3.

The ERG notes that the CS does not include any clinical or economic evidence relating to paediatric patients; the draft SmPC for blinatumomab[24] notes that the safety and efficacy of blinatumomab in paediatric patients have not yet been established. The draft SmPC also states that there is limited experience with blinatumomab in patients ≥75 years of age, and that the safety and efficacy of blinatumomab have not been studied in patients with severe renal impairment or in patients with severe hepatic impairment.[24]

3.2 Intervention

The intervention under appraisal is blinatumomab (Blincyto®). Blinatumomab is a bispecific T-cell engager antibody construct that binds specifically to CD19 expressed on the surface of cells of B-lineage origin and CD3 expressed on the surface of T-cells.[24] Blinatumomab currently holds an EU marketing authorisation for the treatment of adults with Ph- relapsed or refractory (R/R) BCP-ALL. The CS[1] highlights that blinatumomab is the first and only drug indicated specifically for MRD+ BCP-ALL patients in haematological CR.

Blinatumomab is available as a single vial containing 38.5μg of blinatumomab solution. The current list price for a single vial of blinatumomab is £2,017.[25] A simple discount Patient Access Scheme has been approved by the Department of Health; including the discount, the price of blinatumomab is [redacted] per vial.[1]

Within its MRD+ BCP-ALL indication, patients may receive one cycle of induction treatment followed by up to three additional cycles of blinatumomab consolidation treatment. A single cycle of treatment is comprised of 28 days of continuous intravenous (IV) infusion followed by a 14-day treatment-free interval.[24] The draft SmPC states that when considering the use of blinatumomab as a treatment for MRD+ BCP-ALL, detectable MRD (defined as molecular relapse or molecular failure) should be confirmed in a validated assay with minimum sensitivity of 10^{-4}. Clinical testing of MRD, regardless of the choice of technique, should be performed by a qualified laboratory familiar with the technique.[24]

The draft SmPC[24] states that the decision to discontinue blinatumomab temporarily or permanently, as appropriate, should be made in the case of the following severe (Grade 3) or life-threatening (Grade 4) toxicities: cytokine release syndrome; tumour lysis syndrome; neurological toxicity; elevated liver enzymes, and any other clinically relevant toxicities (as determined by the treating physician).

The draft SmPC[24] lists the following special warnings and precautions for use: neurologic events; infections; cytokine release syndrome and infusion reactions; tumour lysis syndrome; neutropenia and febrile neutropenia; elevated liver enzymes; pancreatitis; leukoencephalopathy including progressive multifocal leukoencephalopathy; immunisations; contraception; medication errors and excipients with known effect.

Contraindications to blinatumomab include hypersensitivity to the active substance or to any of the excipients listed in the SmPC and breast-feeding.[24]

3.3 Comparators

The final NICE scope[22] defines two relevant comparators: (i) retreatment with combination chemotherapy, and (ii) monitor for relapse.

The company's review of clinical effectiveness (see CS,[1] Section B2) did not identify any studies which included head-to-head comparisons of blinatumomab versus either of the comparators listed in the final NICE scope.[22] As a consequence, the company's systematic review focusses on a single historical control comparator study that included adult Ph- BCP-ALL patients who have received country-specific standard care treatments (according to the locations in which the study was conducted), achieved a haematological CR, and subsequently had persistent or relapsed MRD. The range of chemotherapy regimens received by patients within the historical control study is not reported, however, the CS refers to *"standardised treatment protocols developed as part of the European Working Group for Acute*

Lymphocytic Leukaemia (EWALL) collaboration."[1] Clinical advice received by the ERG suggests that this should ensure that patients are treated to a similar standard across countries. The company's model uses propensity score methods with average treatment effect on the treated (ATT) weights to adjust the observed data for the standard care group to reflect the characteristics of the blinatumomab study (BLAST). The model assumes that standard care chemotherapy is given according to the UKALL14 trial maintenance regimen;[19] this is comprised of: (i) vincristine (IV, 1.4mg/m^2 once every 13 weeks); (ii) methotrexate (intrathecal, 12.5mg once every 13 weeks); (iii) prednisolone (oral, 60mg/m^2 5 times every 13 weeks); (iv) mercaptopurine (oral, 75mg/m^2 daily), and (v) methotrexate (oral, 20mg/m^2 weekly). This regimen is used only to estimate the costs of chemotherapy; downstream interventions within both treatment groups include allogeneic HSCT (given pre- and/or post-relapse) and salvage chemotherapy using the FLAG-IDA regimen.

As shown in Table 1, the CS does not consider evidence relating to the "monitor for relapse" comparator; this exclusion is based on the argument that it is highly unlikely that MRD+ patients who have a high risk of relapse would solely be monitored for relapse without any treatment. As noted in Section 3.1, the clinical advisors to the ERG suggested that some older and less fit patients will not be able to undergo HSCT or to tolerate chemotherapy, but may be offered blinatumomab for the treatment of persistent MRD positivity. Therefore, monitoring for relapse is a relevant comparator within this patient subgroup and should have been considered in the CS.

3.4 Outcomes

The final NICE scope[22] lists the following outcomes:

- Overall survival (OS)
- Disease-free survival (DFS)
- Relapse-free survival (RFS)
- Minimal residual disease (MRD) response
- Rate of stem cell transplant
- Adverse effects of treatment
- Health-related quality of life (HRQoL)

The CS[1] reports on all of these outcomes for patients receiving blinatumomab within the BLAST study. The reporting of outcomes for the historical control study is restricted to RFS and OS (see CS,[1] Section B.2.9.4). The company's health economic model is based on data from BLAST and the historical control study relating to RFS, OS, HSCT rates and HRQoL.

3.5 Economic analysis

The CS reports the methods and results of a *de novo* model-based health economic analysis to assess the incremental cost-effectiveness of blinatumomab versus standard care chemotherapy for the treatment of adults with MRD+ B-precursor Ph- ALL in CR1. The company's health economic analysis is detailed and critiqued in Chapter 5.

3.6 Subgroups

The final NICE scope[22] does not specify any subgroups of patients with MRD+ BCP-ALL. The company's indirect comparison and health economic analysis are restricted to patients with MRD+ BCP-ALL who are Ph- and in CR1. The company's clinical effectiveness review includes an analysis of RFS and OS outcomes for BLAST patients in CR2 according to MRD response; however, no comparative analyses are presented against other standard care therapies.

3.7 Special considerations

The CS[1] states that there are no equality issues relating to the use of blinatumomab for the treatment of adult MRD+ BCP-ALL patients in haematological CR.

The CS states that blinatumomab is indicated for a rare condition which affects only a very small number of patients (85 patients per year). According to the CS, these patients have a significant unmet medical need and they may gain substantially from access to blinatumomab. The CS goes on to argue that blinatumomab meets many of the criteria for appraisal under the Highly Specialised Technologies (HST) framework and as such, blinatumomab should be evaluated taking into account a wider range of criteria relating to benefits and costs. As blinatumomab has been referred for appraisal under the Technology Appraisal (TA) programme, this issue is not discussed further within this ERG report.

The CS also claims that on the basis of median OS gains derived from the ATT-weighted propensity score analyses, blinatumomab meets NICE's criteria for life-extending treatments given at the end of life.[26] Undiscounted mean OS estimates for blinatumomab and standard care are not used to support this argument, but can be generated using the company's model. Some of the company's economic analyses (e.g. the probabilistic sensitivity analyses) are interpreted based on the assumption that the end of life criteria are met. The ERG notes that due to the use of parametric cure models, median OS and mean OS estimates diverge significantly. The evidence available to determine whether blinatumomab satisfies NICE's end of life criteria is discussed in Chapter 6.

4. CLINICAL EFFECTIVENESS

This chapter presents a review of the clinical effectiveness evidence provided in the CS[1] for blinatumomab for treating patients with MRD+ BCP-ALL. The clinical evidence provided in the CS comprised a systematic review of randomised controlled trials (RCTs) and observational studies for adults with MRD+ BCP-ALL (Appendix D of the CS).

4.1 Critique of the methods of review

4.1.1 Searches

The company performed one clinical effectiveness search to identify all clinical and safety studies of all treatments for adult ALL patients with MRD positivity after treatment (see CS,[1] Appendix D). For the original searches undertaken in May 2017, several electronic bibliographic databases were searched including MEDLINE in Process [via PubMed], EMBASE [host not reported], the Cochrane Database of Systematic Reviews [CDSR, via Wiley Online Library], the Cochrane Central Register of Controlled Trials [CCRCT, via Wiley Online Library], the Database of Abstracts of Reviews of Effectiveness [DARE, via CRD], the NHS Economic Evaluation Database [NHS EED, via CRD] and the Health Technology Assessment database [HTA, via CRD]. Conference proceedings websites (American Society for Blood and Marrow Transplantation [ASBMT], American Society of Clinical Oncology [ASCO], American Society of Hematology [ASH], European Cancer Organisation/European Society for Medical Oncology [ECCO/ESMO], European Hematology Association [EHA], and International Society for Pharmacoeconomics and Outcomes Research [ISPOR]) were searched covering the period from June 2014 until June 2017.

According to the company's clarification response[17] (question A5), two clinical trials registers were searched on the 7th and 8th June 2017 (clinicaltrials.gov and the World Health Organization International Clinical Trials Registry Platform [WHO ICTRP]). Supplementary searches undertaken by the company also included searching unpublished Amgen studies.

For the systematic literature review searches, the company fully reported the search strategies for all the databases searched in Appendix D of the CS. The population terms comprising MeSH and free-text terms for "ALL" were combined with free-text terms for "minimal residual disease". Whilst the company have included most, if not all, of the terms for "minimal residual disease", the ERG is unable to confirm whether applying this will retrieve all ALL studies which include MRD measurement, if for example, these terms are not mentioned in the title and/or abstracts of publications.

The company applied four limits to the search strategies: (i) to exclude paediatric populations; (ii) to include only human studies; (iii) to include English language publications and (iv) to exclude certain publication types (reports, editorials, reviews, news, letters). The ERG recommends limiting the search by applying 'NOT' to exclude animal studies rather than by limiting using the 'Humans' limit function in PubMed, as the former approach is more sensitive.

The application of the English language limit suggests that the search is prone to a language bias, hence the ERG cannot confirm definitively whether any relevant non-English studies of blinatumomab have been excluded from the company's review.

No adverse event (AE) studies were identified from the searches presented in CS Appendix D. In response to a request for clarification from the ERG (see clarification response,[17] question A5), the company confirmed that a systematic search specifically for AEs for blinatumomab was not performed. The primary source of evidence on AEs was the regulatory authorities' documentation i.e. the European Public Assessment Report (EPAR). The ERG considers that the company should have undertaken a separate search for AE studies.

Aside from the issues relating to the implementation of the company's searches, the ERG considers it unlikely that any relevant studies of blinatumomab have been missed the company's searches.

4.1.2 Inclusion criteria

Appendix D of the CS describes the inclusion and exclusion criteria for acceptance into the systematic review (Table 2). One review was undertaken to identify studies of blinatumomab and its comparators (see CS, Appendix D); all studies of any interventional therapy were eligible for inclusion in the company's review.

Table 2: Eligibility criteria for the company's systematic review (reproduced from CS Appendix D Table 72)

	Inclusion criteria	**Exclusion criteria**
Population	Adult ALL patients with MRD positivity after treatment	• Paediatric patients • MRD- ALL patients
Intervention/comparator	Any interventional therapies	None
Outcomes	Clinical effectiveness and safety • OS • RFS • Event-free survival • MRD complete response rate • Duration of MRD response • Duration of haematologic response • Rate of transplant • Mortality following transplant • Treatment-related mortality • Serious adverse events • Grade 3 or 4 AEs (list to be determined based on the most commonly reported) • Discontinuations due to adverse events • Patient-reported outcomes	Non-clinical outcomes, such as those in pharmacodynamics or *in vitro* studies
Study design	• RCTs of at least 10 patients per arm • Single-arm clinical trials of at least 10 patients • Prospective and retrospective observational studies of at least 10 patients	• Case studies and studies evaluating fewer than 10 patients • Letters, narrative reviews, expert opinions, etc.

As stated in the decision problem (see Table 1), the comparator of "monitor for relapse" that was specified in the final NICE scope[22] was not considered in the CS. The clinical advisors to the ERG noted that some older and less fit patients may not be able to receive HSCT or to tolerate chemotherapy; however, they may be able to tolerate blinatumomab. The clinical advisors noted that this population would be small. The ERG considers that the exclusion of this comparator is not appropriate, although no clinical evidence is reported for the use of blinatumomab in this subgroup. It is unclear whether alternative sources (for example, unpublished registry data) may have provided evidence for this comparator.

The included population relates to adult ALL patients with MRD positivity after treatment. Different technologies could be used to define MRD positivity (multicolour flow cytometry to detect abnormal immunophenotypes; real-time quantitative polymerase chain reaction [RT-qPCR] assays to detect clonal rearrangements in Ig heavy chain genes, and/or T-cell receptor [TCR] genes; RT-qPCR assays to detect fusion genes) (CS Section B.1.3). This was not an inclusion criterion, but was recorded for each study. The ERG considers this to be appropriate.

Included outcomes were AEs, patient-reported outcomes (PROs), and the following clinical effectiveness outcomes: OS; RFS; event-free survival (EFS); MRD complete response rate; duration of haematologic response; rate of transplant, and mortality following transplant. These outcomes are consistent with the final NICE scope,[22] with the addition of duration of haematologic response which was not listed in the final NICE scope.

Study designs eligible for inclusion in the company's review included RCTs, single-arm studies and prospective and retrospective observational studies. The ERG considers this to be appropriate. Studies were only included if they had at least 10 patients (or 10 per arm for RCTs). The ERG considers this criterion to be arbitrary and notes that its application could lead to the exclusion of small but relevant studies. However, following a request for clarification from the ERG (see clarification response,[17] question A6) the company confirmed that no relevant studies were excluded for this reason. Included publications were limited by English language, but not location of study.

Study selection was conducted by two independent reviewers, with disagreements resolved by a third reviewer, in accordance with good practice for systematic reviews.

4.1.3 Critique of data extraction

The company's clarification response[17] (question A7) states that data extraction was conducted independently by two reviewers, with disputes resolved by a third reviewer. The ERG considers this to reflect good practice in systematic reviews.

Based on the information provided in Section B.2 of the CS, it was apparent that relevant data were extracted on study methodology and patient characteristics. CS Appendix D explicitly states that data were extracted on definition of MRD and subgroups according to CR status after first-line or salvage treatment (see CS, Table 73).

Data extracted for the three studies and included in the CS (see Section 4.2) were checked by the ERG against clinical study reports (CSRs) and were found to be accurate.

4.1.4 Quality assessment of included studies

The company's quality assessment was conducted independently by two reviewers, with disputes resolved by a third reviewer (see clarification response,[17] question A7), as is good practice in systematic reviews.[27]

Quality assessment of the two included blinatumomab studies (BLAST and MT103-202), and the retrospective study of standard care chemotherapy (Study 20120148), was presented in CS Appendix D (see CS,[1] Table 74). The quality assessment tool used in the CS was the Risk Of Bias In Non-randomised Studies of Interventions (ROBINS-I) checklist.[28] This checklist is designed to assess risk of bias within a given study, but does not address external validity, and therefore does not address issues relating to generalisability or the limitations of particular study designs.[28] The three included studies were all single-arm, open-label studies; the CS acknowledges that non-randomised study designs are subject to limitations. Based on the ROBINS-I checklist, the company deemed the overall risk of bias of all three studies to be low.

The ROBINS-I checklist is designed for non-randomised studies of interventions *"that compare the health effects of two or more interventions"* (detailed guidance is available from https://sites.google.com/site/riskofbiastool//welcome/home). Whilst there is no universally accepted validated tool for critically appraising single-arm studies, several checklists have been developed and applied to case series.[29] The ERG assessed the quality of the single-arm studies based on the criteria for case series suggested by the Centre for Reviews and Dissemination (CRD, see Table 3).[30,29]

Table 3: Quality assessment of the three included studies

CRD criteria	BLAST[31]	MT103-202[32]	Study 20120148[33]
Is the study based on a representative sample selected from a relevant population?	Yes	Yes	Yes
Are the criteria for inclusion explicit?	Yes	Yes	Yes
Did all individuals enter the survey at a similar point in their disease progression?	Yes	Yes	Yes
Was follow-up long enough for important events to occur?	Yes	Yes	Yes
Were outcomes assessed using objective criteria or was blinding used?	Outcome assessors were not blinded OS – objective criteria MRD response– objective criteria RFS – objective criteria HRQoL – at risk of bias	Outcome assessors were not blinded MRD – objective criteria RFS – objective criteria	Outcome assessors were not blinded OS – objective criteria RFS – objective criteria

The studies were well conducted according to CRD criteria.[30] Prognostic factors such as disease stage and age were reported in the CS for all three included studies. Statistical analyses including subgroup analyses were pre-specified.[31-33] In the BLAST study there was a low risk of attrition bias, all patients included at baseline and treated with at least one cycle of blinatumomab were included in the OS analysis, and the majority of these patients were included in the RFS (95%) and MRD (97%) analyses. MT103-202 included all 20 patients treated with at least one cycle of blinatumomab in MRD and RFS analyses.

Single-arm studies are low on the hierarchy of study quality as they are associated with potential biases.[34] The absence of blinding leads to a risk of performance bias.[27, 30] The lack of randomisation leads to a risk of selection bias.[27, 30]

Eligibility criteria for all three included studies were adequately described in the CS. However, it was not clear from the CS how patients were identified for recruitment into the blinatumomab studies and whether patients were recruited consecutively.[29, 35] The company's clarification response[17] (question A16) provides reasons for not enrolling 95 screened patients, most of which were due to patients not meeting eligibility criteria for MRD level ($< 1 \times 10^{-3}$) or having an overt relapse.

Single-arm studies also have a risk of detection bias due to the absence of blinding. One means by which the risk of bias can be reduced in open-label studies is to introduce blinded outcome assessors. However, in the included studies, the lack of blinding is unlikely to impact on OS, MRD or RFS. The HRQoL outcome is necessarily prone to bias as it is patient-reported and therefore assessor-blinding is not possible in an open-label study.[36]

Study 20120148 comprised a retrospective analysis of existing clinical databases. Retrospective studies are more likely to be susceptible to bias than prospective studies, particularly selection bias.[30, 37] However, Study 20120148 was used in the CS to select a population comparable to that of the BLAST study, rather than to provide a population representative of all MRD+ BCP-ALL patients.

4.1.5 Evidence synthesis

Due to the lack of RCT data to inform the comparative effectiveness of blinatumomab versus standard care chemotherapy, the company synthesised data from BLAST and Study 20120148 using indirect comparison methods. Further details of this analysis are provided in Section 4.4.

4.2 Included blinatumomab studies

The CS[1] included three studies identified by the systematic searches. Two studies were of blinatumomab (MT103-202 and BLAST). The third study was a historical comparator study (Study 20120148, described in Section 4.3). All three studies were sponsored by Amgen and information was provided in CSRs[31,32,33] and the BLAST protocol.[38] At the time of writing, BLAST was published in two abstracts.[39, 40]

No relevant RCTs were identified by the company or by the ERG. The ERG does not believe that any relevant studies of blinatumomab retrieved from the searches were excluded from the CS.

An ERG search of the U.S. National Institutes of Health (NIH) clinical trials registry[41] identified two potentially relevant ongoing studies, however, at the time of writing, the completion dates for these studies were more than 12 months in the future. Study NCT02458014 (Blinatumomab in Patients with B-cell Lineage Acute Lymphocytic Leukaemia with Positive Minimal Residual Disease) has an estimated primary completion date of September 2020. Study NCT02767934 (Pembrolizumab in Treating Minimal Residual Disease in Patients with Acute Lymphoblastic Leukaemia), which includes B-cell as well as T-cell ALL, has an estimated primary completion date of January 2019.

4.2.1 Study characteristics of blinatumomab studies

The two included studies of blinatumomab (MT103-202[32] and BLAST[31]) were both single-arm studies. Study characteristics are shown in Table 4. Within the limitations of the study design, the studies were well conducted, however, as noted in Section 4.1.4, there are biases associated with single-arm studies.

Table 4: Characteristics of included blinatumomab studies

Study	Reference(s)	Study design	Population	Number enrolled	Intervention	Primary outcome	Dates of enrolment	Follow-up
MT103-202 NCT00560794	Amgen CSR 2013[32]	Phase II, single-arm, open-label, multicentre	Adult MRD+ BCP-ALL patients in haematological CR after front-line therapy	[redacted] 20 received at least one cycle and included in efficacy analysis (of 21 who received at least one infusion and were included in safety analysis)	Blinatumomab 15µg/m^2/day continuous infusion. Up to 10 cycles	Incidence of MRD negativity/response** within 4 cycles of treatment with blinatumomab	2008 – 2009	Primary efficacy 24 weeks (4 cycles) Safety up to 4 weeks after last treatment
BLAST MT103-203 NCT01207388	Amgen CSR 2016[31] Amgen protocol 2010[38] Goekbuget 2014[39]	Phase II, single-arm, open-label, international, multicentre	Adult MRD+ BCP-ALL patients in haematological CR after front-line therapy	[redacted], 116 received at least one infusion	Blinatumomab 15µg/m^2/day continuous infusion* Up to 4 cycles	Proportion of patients who achieve complete MRD response defined by absence of MRD after one cycle of treatment	2010 – 2014	Safety 30-days Efficacy 9, 12, 18, and 24 months Survival 30, 36, 42, 48, 54, 60 months

Information from CS Section B.2.2 and B.2.3 and Appendix D, CSRs,[31, 32] and U.S. National Institutes of Health clinical trials registry[41]
** one cycle = continuous infusion for four weeks followed by two-week infusion-free interval*
*** MRD negativity/response defined as bcr/abl and/or t[4;11] below detection limit and/or individual rearrangements of immunoglobulin or TCR genes below 10^{-4} [31] bcr/abl = "breakpoint cluster region/gene on human chromosome #9"* [31]

Eligibility criteria

BLAST inclusion criteria

Patients were eligible for inclusion in the study only if all the following criteria applied:

- Patients with BCP-ALL in complete haematological remission defined as less than 5% blasts in bone marrow after at least three intense chemotherapy blocks (e.g., GMALL induction I–II/consolidation I, induction/intensification/consolidation or three blocks of Hyper CVAD)
- Presence of MRD at a level of $\geq 10^{-3}$ (molecular failure or molecular relapse) in an assay with a sensitivity and a lower level of quantification of 10^{-4} documented after an interval of at least 2 weeks from last systemic chemotherapy
- For evaluation of MRD, patients must have had at least one molecular marker based on individual rearrangements of immunoglobulin or TCR-genes or a flow cytometric marker profile evaluated by a national or local reference lab approved by the sponsor
- Bone marrow specimen from primary diagnosis (enough DNA [30pg] or a respective amount of cell material) for clone-specific MRD assessment must have been received by central MRD lab and lab must have confirm that the sample is available
- Bone marrow function as defined below:
    - ANC (Neutrophils) ≥1,000/µL
    - Platelets ≥50,000/µL (transfusion permitted)
    - Haemoglobin (HB) level ≥9g/dI (transfusion permitted)
- Renal and hepatic function as defined below:
    - AST (GOT), ALT (GPT), and AP <2 x ULN
    - Total bilirubin <1.5 x ULN
    - Creatinine clearance ≥50 mL/min (calculated e.g. per Cockroft & Gault)
- Negative HIV test, negative hepatitis B (HbsAg) and hepatitis C virus (anti-HCV) test
- Negative pregnancy test in women of childbearing potential
- Eastern Cooperative Oncology Group (ECOG) Performance Status 0 or 1
- Age ≥18 years
- Ability to understand and willingness to sign a written informed consent
- Signed and dated written informed consent.

BLAST exclusion criteria

Patients were excluded from participation in the study if any of the following criteria applied:

- Presence of circulating blasts or current extra-medullary involvement by ALL
- History of relevant central nervous system (CNS) pathology or current relevant CNS pathology (e.g. seizure, paresis, aphasia, cerebrovascular ischemia/haemorrhage, severe brain injuries,

dementia, Parkinson's disease, cerebellar disease, organic brain syndrome, psychosis, coordination or movement disorder)

- Current infiltration of cerebrospinal fluid (CSF) by ALL
- History of or active relevant autoimmune disease
- Prior allogeneic HSCT
- Eligibility for treatment with TKIs (i.e., Ph+ patients with no documented treatment failure of or intolerance/contraindication to at least 2 TKIs)
- Systemic cancer chemotherapy within 2 weeks prior to study treatment (except for intrathecal prophylaxis)
- Radiotherapy within 4 weeks prior to study treatment
- Autologous HSCT within six weeks prior to study treatment
- Therapy with monoclonal antibodies (rituximab, alemtuzumab) within 4 weeks prior to study treatment
- Treatment with any investigational product within four weeks prior to study treatment
- Previous treatment with blinatumomab
- Known hypersensitivity to immunoglobulins or to any other component of the study drug formulation
- History of malignancy other than ALL within five years prior to treatment start with blinatumomab, except for basal cell or squamous cell carcinoma of the skin, or carcinoma "in situ" of the cervix
- Active infection, any other concurrent disease or medical condition that are deemed to interfere with the conduct of the study as judged by the investigator
- Nursing women or women of childbearing potential not willing to use an effective form of contraception during participation in the study and at least 3 months thereafter or male patients not willing to ensure effective contraception during participation in the study and at least three months thereafter.

Study MT103-202 eligibility criteria (from CS[1] Section B.2.3 and the NIH clinical trials registry[41]) were as follows:

Study MT103-202 inclusion criteria

- Adults (≥18 years of age) with BCP-ALL
- MRD positivity at a level of at least $1x10^{-4}$ at any point after the first consolidation chemotherapy block of front-line therapy
- ECOG Performance Status < 2.

Study MT103-202 exclusion criteria

- Current extramedullary involvement
- History of (or current) clinically relevant CNS pathology or autoimmune disease
- Prior autologous HSCT (within 6 weeks) or allogeneic HSCT (at any time)
- Chemotherapy or radiotherapy (within 4 weeks)
- Therapy with monoclonal antibodies (within 6 weeks)
- Known hypersensitivity to immunoglobulins or to any other component of the study drug formulation
- History of malignancy other than ALL within five years prior to treatment start with blinatumomab, except for basal cell or squamous cell carcinoma of the skin, or carcinoma "in situ" of the cervix
- Active infection, any other concurrent disease or medical condition that are deemed to interfere with the conduct of the study as judged by the investigator
- Nursing women or women of childbearing potential not willing to use an effective form of contraception during participation in the study and at least 3 months thereafter or male patients not willing to ensure effective contraception during participation in the study and at least three months thereafter.

Table 5 presents the baseline characteristics of BLAST and MT103-202. The studies had similar baseline ages (BLAST median age=45 years, MT103-202 mean age=50 years). The majority of participants were Ph-, with n=[redacted] Ph+ participants in each study. In the BLAST study, the majority of patients (65%) were in first CR. The demographics of the study were considered by clinical advice to be similar to those of the UK population with BCP-ALL who have MRD activity while in remission. In the BLAST study, the majority of patients (84%) had a baseline MRD level of between 10^{-3} and 10^{-1}, where patients are classed as MRD+ with disease measurable to 10^{-4}. These MRD levels may not necessarily reflect those of the UK population, but reflect the eligibility criteria for the blinatumomab studies.

Table 5: Baseline characteristics of participants in BLAST and MT103-202

Commented [LM1]: CHECK ACIC

Baseline characteristic	BLAST MT103-203 (n=116)		MT103-202 (Pilot) (n=20)	
Male sex, n (%)	68 (59)		8 (40.0)	
Age, years	Median (range)	45.0 (18–76)	Mean (SD)	49.8 (18.3)
Age, n (%)	≥18 to <35 years	36 (31.0)	20-30 years	3 (15.0)
	≥35 to <55 years	41 (35.3)	31-40 years	5 (25.0)
	≥55 to <65 years	24 (20.7)	41-50 years	2 (10.0)
	≥65 years	15 (12.9)	51-60 years	1 (5.0)
			61-70 years	7 (35.0)
			> 70 years	2 (10.0)
Median time from prior treatment (range), months	[redacted]		NR	
Relapse history, n (%)	First CR	75 (65)	NR	
	Second CR	39 (34)		
	Third CR	2 (2)		
Baseline MRD levels at cetral laboratory, n (%)	$\geq 10^{-1}$ <1	9 (7.8)	NR	
	$\geq 10^{-2}$ $<10^{-1}$	45 (38.8)		
	$\geq 10^{-3}$ $<10^{-2}$	52 (44.8)		
	$<10^{-3}$	3 (2.6)		
	Below LLQ	5 (4.3)		
	Unknown	2 (1.7)		
Philadelphia chromosome disease status, n (%)	Positive	5 (4.3)	Positive	[redacted] (CS Clarification response[17] A2)
	Negative	111 (95.7)	Negative	[redacted]
Ethnicity, n (%)	White: 102 (87.9) Asian: 1 (0.9) Mixed: 1 (0.9) Unknown: 12 (1.3)		Caucasian 20 (100.0)	
Genetic alterations, n (%)	Confirmed t(4;11) Translocation / MLL-AF4+	5 (6.8)	Confirmed t(4;11) Translocation / MLL-AF4+	2 (10.0)
			bcr/abl above detection limit (all)	5 (25.0)
WBC at diagnosis (>30,000/mm^3), n (%)	18 (15.5)		NR	

Information from CS Section B.2.3, Goekbuget 2014[39] and U.S. National Institutes of Health clinical trials registry[41]
Values in parentheses represent percentages
CR - complete response; LLQ - lower limit of quantification; WBC - white blood cell; NR - not reported

Prior ALL treatment received by patients in the BLAST study was provided in the company's clarification response[17] (question A12), and is shown in Table 6. Prior front-line treatment had been received by [redacted] of the patients, [redacted] had received treatment for first relapse, and only [redacted] had received treatment for second relapse. For prior anti-tumour drug treatment, [redacted]

[redacted] Clinical advice received by the ERG suggested that these treatments would lead to most patients in the study having a similar level of disease to those patients seen in current practice who are eligible for treatment using blinatumomab in England. Although PETHEMA is quite different from practice in England, only a small percentage of patients ([redacted]) received this regimen.

Table 6: BLAST study Prior ALL treatment (reproduced from company's clarification response question A12)

Characteristic Category	Full analysis set (n=116)
Maximum line of therapy	
Front-line treatment	[redacted]
First relapse treatment	[redacted]
Second relapse treatment	[redacted]
Front-line treatment	[redacted]
Pre-phase	[redacted]
GMALL	[redacted]
combination of regimen /other	[redacted]
GMALL elderly	[redacted]
GRAALL	[redacted]
UKALL	[redacted]
GIMEMA	[redacted]
PETHEMA	[redacted]
FLAG-Ida	[redacted]
NILG	[redacted]
TKI	[redacted]
FRAALLE	[redacted]
Hyper-CVAD	[redacted]
iBFM	[redacted]
AIEOP	[redacted]
HOVON	[redacted]
ALL-2009	[redacted]
ALL-2009 elderly	[redacted]
EWALL elderly	[redacted]
GRAAPH	[redacted]
LALA94	[redacted]
Romanian Group for ALL	[redacted]

Concomitant medications allowed are shown in Table 7. Clinical advice received by the ERG suggested that these are similar to current practice in England.

Table 7: Concomitant medications allowed in the blinatumomab studies (data extracted from CS Table 12)

	BLAST	**MT103-202**
Permitted medications	Prior to the start of cycle 1: CSF (cerebrospinal fluid) prophylaxis A corticosteroid Prior to the start of subsequent cycles: A corticosteroid During the treatment period: Dexamethasone in the case of neurologic events Following treatment cycles 2 and 4 immediately after bone marrow aspiration: CSF prophylaxis After completion of study treatment for patients who did not undergo HSCT: CSF prophylaxis Patients at high risk for CMV infection: Intensive CMV-PCR follow-up or prophylactic CMV treatment	Premedication for each treatment cycle included a corticosteroid to suppress cytokine release (100mg methylprednisolone IV at 1 hour prior to start of blinatumomab infusion or prior to restart if infusion interruption > 12 hours) and thrombosis prophylaxis by low molecular weight heparin (subcutaneous) during the first 7 days of each treatment cycle CNS prophylaxis was administered with the following intrathecal triple combination regimen at absolute doses: dexamethasone 4mg, methotrexate 15mg, cytosine-arabinoside 40mg. If the patient had MRD response after cycle 1 of treatment, the triple combination regiment was administered immediately after the first bone marrow aspiration study on day 28 of cycle 2 In non-responders, after cycle 1 demonstrated detectable MRD, the triple combination regimen was administered after cycle 3 of treatment immediately after bone marrow aspiration on cycle day 28 of cycle 3. CNS prophylaxis continued every 3 months Small molecule TKIs registered for the treatment of ALL disease were permitted as concomitant treatment of patients with bcr/abl positive MRD if the patients developed MRD relapse on TKIs or whose MRD persisted on TKIs for more than 8 weeks For symptomatic treatment of fever, metamizole was administered
Disallowed medications	Any anti-tumour therapy Any other investigational agent Chronic systemic high-dose corticosteroid therapy Any other immunosuppressive therapies Non-steroidal anti-inflammatory drugs Paracetamol/acetaminophen was allowed TKIs	Any anti-tumour therapy other than blinatumomab as indicated in the protocol Any other investigational agent Chronic systemic high-dose corticosteroid therapy Other immunosuppressive therapies Stem-cell transplantation Any use of NSAIDs (nonsteroidal anti-inflammatory drugs) except for paracetamol

4.2.2 Clinical effectiveness in the blinatumomab studies

Overall Survival

Study MT103-202 did not include OS as an outcome measure. OS data from BLAST are shown in Table 8 and Figure 2. OS was defined as the time from first blinatumomab treatment until death due to any cause, with patients who did not die being censored at their last contact date (see CS,[1] Section B.2.6.1).

Results from Cox proportional hazards models including treatment as a covariate in the model are presented by the company. Results from the primary analysis, described by the company as "without censoring at HSCT" are used for the health economic model. Results are also presented including a time-dependent covariate for HSCT, described by the company as "with censoring at HSCT"; the company states that this analysis was conducted to account for differences between transplant rates in BLAST and the historical control, hence it better isolates the blinatumomab treatment effect not affected by use of transplant.[1] The primary analysis (without censoring at HSCT) is considered by the ERG to be most appropriate as, according to the CS, blinatumomab is not intended to displace HSCT in the treatment pathway.

After a median follow-up of 18 months, median OS was [redacted]

Figure 2: OS in BLAST (reproduced from CS Figure 13)

[redacted]

Subgroup analysis of OS outcomes was conducted according to the following factors: age; gender; Philadelphia status; patients by t(4;11) translocation and/or MLLAF4+ ALL haematological remission; risk stratification; relapse history; MRD level at baseline; white blood cells (WBC) at first diagnosis; chemo-resistance after the first week of chemotherapy; need of salvage therapy for CR; previous anti-tumour radiotherapies; incidence of neurologic events during cycle 1; time from diagnosis to start of blinatumomab; time from last treatment to start of blinatumomab, and clinical trial material from manufacturing process 4/5. The only subgroup which was found to differ significantly for OS was Ph status, with Ph- patients experiencing a significantly [redacted] median OS than Ph+ patients ([redacted]) (see CS,[1] Section B2.7). This was based on only 5 Ph+ patients, all of whom were in CR2/3 rather than CR1, hence the ERG considers that the interpretation of this subgroup finding should be treated with caution.

Table 8: Summary of OS outcomes in BLAST

	BLAST (n=116)	
Outcome	**OS not censored at HCST**	**OS censored at HCST**
Events, n (%)	[redacted]	[redacted]
Censors, n (%)	[redacted]	[redacted]
OS % (18 months)	[redacted]	[redacted]
95% CI	[redacted]	[redacted]
Median (months)	[redacted]	[redacted]

Information from CS Section B.2.6.1 (Table 21 of the CS) and U.S. National Institutes of Health clinical trials registry[41]
n.e. = not estimable

Relapse-free survival

Haematological RFS in BLAST was measured from the first dose of blinatumomab until the first assessment of documented relapse (either haematological (>5% leukaemia cells in bone marrow as measured by cytological, microscopic assessment, presence of circulating leukaemia blasts) or extramedullary leukaemia), secondary leukaemia, or death due to any cause.[41] In the MT103-202 study, time to haematological relapse was defined as the time between the start of first infusion of blinatumomab and the first result of haematological relapse, defined as >5% leukaemia cells in bone marrow.[41]

RFS data were provided by 110 patients in the BLAST study who were in haematological CR at baseline, excluding Ph+ participants (see Table 9) as Ph+ patients were excluded from the pre-specified secondary analyses.[41] At 18-months follow-up, the uncensored median time to haematological relapse was [redacted]. In the MT103-202 study, RFS was [redacted] at five years.

Figure 3: RFS in BLAST (reproduced from CS Figure 10)

[redacted]

Subgroup analysis of RFS outcomes was conducted according to the following factors: age; gender; Philadelphia status; patients by t(4;11) translocation and/or MLLAF4+ ALL haematological remission; risk stratification; relapse history; MRD level at baseline; WBC at first diagnosis; chemo-resistance after the first week of chemotherapy; need of salvage therapy for CR; previous anti-tumour radiotherapies; incidence of neurologic events during cycle 1; time from diagnosis to start of blinatumomab; time from last treatment to start of blinatumomab, and clinical trial material from manufacturing process 4 or 5. The only subgroup found to differ significantly was relapse history. [redacted] (see CS,[1] Section B2.7).

[redacted]

Table 9: Summary of RFS outcomes in BLAST and MT103-202

Study	**BLAST n=110**		**MT103-202 (Pilot) n=20**
Outcome	**RFS not censored at HCST**	**RFS censored at HCST**	
Events, n (%)	[redacted]	[redacted]	NR
Censors, n (%)	[redacted]	[redacted]	NR
RFS %	[redacted]	[redacted]	[redacted]
95% CI	[redacted]	[redacted]	
Median RFS	[redacted]	[redacted]	[redacted]

(months)			

Information from CS Section B.2.6.1 (Table 20 and Figure 10 of the CS) and CS Section B.2.6.2

Minimal residual disease response

Within the BLAST study, complete MRD response was defined as no polymerase chain reaction (PCR) amplification of individual rearrangements of immunoglobulin (Ig)- or TCR -genes detected (the minimum required sensitivity of 1 x 10^{-4}) after completion of the first cycle (see CS Section B.2.3 and US NIH clinical trials registry[41]). [redacted][38]

For patients in the MT103-202 study, the primary endpoint was MRD response rate within four cycles of blinatumomab. For Ph+ or translocation (t) (4;11) patients, response was achieved when Ph or t(4;11) was below detection limit and individual rearrangements of immunoglobulin or TCR genes were below 1 x 10^{-4}. For Ph- and t(4;11) negative, response was achieved when individual rearrangements of immunoglobulin or TCR genes were below 1 x 10^{-4} (see CS,[1] Section B.2.3 and US NIH clinical trials registry[41]). MRD response outcomes are presented in Table 10. All 80% of patients achieving MRD response did so within one cycle.

In the BLAST study, three patients were excluded from the MRD response analysis due to missing data (n=1) or assays with a sensitivity of 5×10^{-4} (n=2).[39] Data on MRD response from 113 patients in BLAST are shown in Table 10. A total of ninety patients ([redacted]) achieved MRD response after one or more cycles of blinatumomab treatment, with 88 of these patients responding within one cycle.[39] There was a higher rate of response for patients in first CR 82% (95% CI 72% to 90%), than in second CR 71% (95% CI 54% to 85%) or third CR 50% (95% CI 1% to 99%); however, only two patients were in third CR (see Table 5), hence results on this subgroup should be treated with caution.[39] For other subgroups, there was no significant difference in MRD response (age, sex, line of treatment, and MRD levels).[39]

Table 10: MRD response in the blinatumomab studies

Outcome	BLAST [redacted]	MT103-202 (Pilot) (n=20)
Complete MRD response after 1 cycle n (%,95% CI)	[redacted]	[redacted]
Complete MRD response after ≥1 cycle n (%, 95% CI)	[redacted]	[redacted]
Median time to MRD response, days	[redacted]	NR
Duration of median response, months (uncensored)	[redacted]	[redacted]).
Duration of median response, months (censored at HCST or post-blinatumomab chemotherapy)	[redacted]	NR

Information from CS Section B.2.6.1 (Tables 18 and 19 of the CS) and CS Section B.2.6.2 and Goekbuget 2014.[39]
**participants who were in haematological complete remission at treatment start, excluding Ph+, who had an MRD complete response at cycle 1*

Rate of stem cell transplant

For Study MT103-202 [redacted] patients underwent HCST (see CS,[1] Section B.2.6.2). In BLAST, [redacted] [redacted] patients underwent HCST, of whom [redacted] were in complete haematological CR at the time of HSCT. Within the group of 74 Ph- patients who underwent HSCT prior to relapse, the 100-day mortality probability was 7%.[1]

Health-related quality of life

BLAST measured HRQoL using the European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire C30 (EORTC QLQ-C30),[42] a validated, cancer-specific patient reported outcome questionnaire, and the Euroqol 5-Dimensions questionnaire (EQ-5D, available from: https://euroqol.org/), a standardised measure of generic health status. HRQoL results from BLAST are shown in Table 11 and Table 12, respectively. [redacted] HRQoL as measured by EQ-5D did not change significantly from baseline to the end of the core study.

Table 11: Change from baseline in EORTC-QLQ-C30 scales in BLAST (n=116) (reproduced from CS Table 24)

EORTC-QLQ-C30 Scale	Baseline, mean (SE) (Max=100)	Greatest change from baseline in cycles 1 to 4, mean (SE)/cycle	Change from baseline at end of core study, mean (SE)
Global health status	[redacted]	[redacted]	3.9 (2.4)
Physical function	[redacted]	[redacted]	2.2 (1.9)
Role functioning	[redacted]	[redacted]	1.4 (3.5)
Emotional functioning	[redacted]	[redacted]	5.3 (2.7)
Cognitive functioning	[redacted]	[redacted]	-2.3 (2.5)
Social functioning	[redacted]	[redacted]	14.9 (3.8)
Fatigue	[redacted]	[redacted]	-5.4 (2.4)
Nausea and vomiting	[redacted]	[redacted]	-2.3 (2.0)
Pain	[redacted]	[redacted]	-1.4 (2.7)
Dyspnoea	[redacted]	[redacted]	-0.9 (2.9)
Insomnia	[redacted]	[redacted]	3.7 (3.5)
Appetite loss	[redacted]	[redacted]	-9.1 (3.4)
Constipation	[redacted]	[redacted]	0 (2.2)
Diarrhoea	[redacted]	[redacted]	0.0 (2.3)
Financial difficulties	[redacted]	[redacted]	-0.9 (2.9)

EORTC-QLQ-C30 - European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire-C30; SE - standard error

Table 12: Change from baseline in EQ-5D domains in BLAST (n=116) (reproduced from CS Table 25)

EQ-5D scale	Baseline, mean (SE)	Greatest change from baseline in cycles 1 to 4, mean (SE)/cycle	Change from baseline at end of core study, mean (SE)
Mobility	[redacted]	[redacted]	0 (0.1)
Self-care	[redacted]	[redacted]	0 (0.0)
Usual activity	[redacted]	[redacted]	-0.1 (0.1)
Pain/discomfort	[redacted]	[redacted]	-0.1 (0.1)
Anxiety/depression	[redacted]	[redacted]	-0.1 (0.1)

EQ-5D - EuroQol 5-dimensions; SE - standard error

Adverse events

Differences in treatment regimen for BLAST and MT103-202

The CS[1] (Section B.2.10.1) reports both treatment-emergent and treatment-related (considered to be related to blinatumomab) AEs for MRD+ BCP- ALL patients. Pooled data from the BLAST study[31] (n=116) and MT103-202[32] (n=21) are reported in the CS.[1] A meta-analysis of data from the two studies was not conducted. The CSR for BLAST[31] reports a median treatment duration of 55 days, whilst the CSR for MT103-202[32] reports a median treatment duration of 87.3 days. The CSRs report that the timeframe for recording of AEs was from the first dose of blinatumomab to 30 days after the last dose for BLAST, and from the first dose of blinatumomab to 4 weeks after the last dose for MT103-202. For BLAST, the CSR reports a dosing regimen of blinatumomab continuous IV infusion at 15µg/m^2/day at a constant flow rate over 28 days, followed by an infusion-free interval of 14 days, for up to 4 cycles.[31] For MT103-202, the CSR reports a dosing regimen of blinatumomab continuous IV infusion at 15µg/m^2/day at a constant flow rate over 28 days, followed by an infusion-free interval of 14 days, for up to 7 cycles in patients who showed neither MRD progression nor response.[32] Patients who had achieved MRD response were administered 3 additional cycles of treatment, up to a maximum of 10 cycles. For MT103-202, a dose increase to 30µg/m^2/day was permitted where there was evidence of insufficient response.[32]

Numbers of adverse events (BLAST and MT103-202)

A summary of the pooled treatment-emergent and treatment-related AEs for MRD+ BCP-ALL patients from BLAST and MT103-202 is presented in Table 13. Amongst MRD+ BCP-ALL patients, [redacted] participants experienced at least one treatment-emergent AE. [redacted] of participants experienced an AE classed as serious; [redacted] of patients had Grade ≥3 AEs; and [redacted] of patients had Grade ≥4 AEs. [redacted] participants experienced treatment-emergent AEs that led to the discontinuation of treatment; [redacted] of participants had a serious adverse event (SAE); [redacted] of patients had Grade ≥3 AEs; and [redacted] of patients had Grade ≥4 AEs. [redacted] suffered AEs that were considered to be treatment-related. [redacted] of participants experienced a treatment-related AE that was classed as serious; [redacted] that were classed as Grade ≥3; and [redacted] that were classed as Grade ≥4. [redacted] participants [redacted] experienced treatment-related AEs that led to the permanent discontinuation of treatment. [redacted] of participants had a serious event; [redacted] Grade ≥3; [redacted] Grade ≥4.

The most common treatment-emergent events of interests (EOIs) were [redacted] of participants); [redacted], and [redacted]. Other EOIs with a frequency of ≥5% were [redacted]; [redacted]; [redacted], [redacted]; and [redacted].

There was [redacted] which was considered related to blinatumomab, and [redacted] which was not considered to be treatment-related.

A summary of disaggregated data reporting the frequency of SAEs for BLAST and MT103-202 is presented in Table 14. These data were taken from the US NIH clinical trials registry and cross-referenced against both CSRs (BLAST CSR Table 14-6.25; MT103-202 Table 12-4). [redacted] of patients in BLAST and [redacted] of patients in MT103-202 experienced an SAE. The most common of these were blood and lymphatic system disorders, infections and infestations, injury, poisoning or procedural complications, and nervous system disorders. In BLAST,[31] [redacted] patients experienced SAEs classified as general disorders. The most common of these was pyrexia. No reports of SAEs classified as general disorders are reported in the CSR for MT103-202.[32] In addition, in BLAST, [redacted] patients experienced SAEs relating to investigations, whilst none were reported in MT103-202.

The CS[1] draws comparisons between the pooled data from BLAST and MT103-202 and the known safety profile of blinatumomab in adult patients with relapsed or refractory Ph-BCP-ALL. This profile comprised pooled data from MT103-206, MT103-211, and TOWER. Table 15 presents data for AEs for these two sets of pooled data, as reported in the CS. Safety profiles are consistent between these populations, with the following exceptions: (i) treatment-emergent Grade ≥3 AEs were lower in MRD+BCP-ALL patients ([redacted]); (ii) there was a higher rate of treatment-related AEs for the MRD+BCP-ALL population ([redacted]), and (iii) there was a difference for treatment-related SAEs, which were higher in MRD+BCP-ALL patients ([redacted]), although the CS reports that this is likely due to a high rate of Grade ≥2 AEs. Grades ≥3 and 4 AEs were comparable between the two populations ([redacted]). With respect to treatment-emergent EOIs, whilst there was a comparable rate of any-grade EOIs between populations, a lower rate of EOIs Grade ≥3 and Grade ≥4 was reported in the MRD+BCP-ALL population ([redacted], respectively). Both populations experienced similar rates of neurological AEs ([redacted]). A lower rate of cytokine release syndrome was reported in MRD+BCP-ALL patients compared with Ph-BCP-ALL patients ([redacted]). The CS suggests this may be a result of a lower disease burden in the MRD+BCP-ALL population in haematological CR. Rates of treatment interruptions were consistent between both populations.

The CS[1] reports that the safety profile of blinatumomab in adult MRD+ BCP-ALL patients reflects its known safety profile in a Ph-BCP-ALL population, with no new risks suggested. The blinatumomab EPAR[43] including AE data for the MRD+BCP-ALL population was unavailable at the time of writing (EMA accessed 30th January 2018), but is expected to be published early 2018 (see CS,[1] Appendix C).

Table 13: Incidence of treatment-emergent and treatment-related AEs from pooled data from the BLAST study and MT103-202 for MRD+ BCP-ALL (adapted from CS Table 30)

Event	Treatment-emergent AEs [redacted]	Treatment-related AEs [redacted]
All AEs, n (%)	[redacted]	[redacted]
Serious	[redacted]	[redacted]
Grade ≥3	[redacted]	[redacted]
Grade ≥4	[redacted]	[redacted]
Fatal*	[redacted]	[redacted]
Leading to permanent discontinuation of blinatumomab	[redacted]	[redacted]
Serious	[redacted]	[redacted]
Grade ≥3	[redacted]	[redacted]
Grade ≥4	[redacted]	[redacted]
Fatal*	[redacted]	[redacted]
Leading to interruption of blinatumomab	[redacted]	[redacted]
Serious	[redacted]	[redacted]
Grade ≥3	[redacted]	[redacted]
Grade ≥4	[redacted]	[redacted]
Fatal*	[redacted]	[redacted]

** Fatal events that occurred within 30 days of last blinatumomab treatment*

Table 14: SAEs in MRD+ BCP-ALL patients for BLAST and MT103-202

SAE	BLAST MT103-203 (n=116)		MT103-202 (Pilot) (n=21)	
SAEs	73		10	
Blood and lymphatic system disorders	[redacted]	Anaemia (1); bone marrow failure (1); febrile neutropenia (2); leukopenia (1); neutropenia (5); thrombocytopenia (1)	[redacted]	Leukopenia (1); lymphopenia (6); thrombocytopenia (1)
Cardiac disorders	[redacted]	Sinus bradycardia (1); sinus tachycardia (1)	[redacted]	NR
Gastrointestinal disorders	[redacted]	Abdominal pain (1); diarrhoea (1); gastrointestinal haemorrhage (1)	[redacted]	NR
General disorders	[redacted]	Device issue (1); device malfunction (2); fatigue (1); gait disturbance (1); infusion site extravasation (1); product contamination microbial (1); puncture site pain (1); pyrexia (17); thrombosis in device (1)	[redacted]	NR
Hepatobiliary disorders	[redacted]	Hepatotoxicity (1)	[redacted]	NR
Immune system disorders	[redacted]	Cytokine release syndrome (2); hypersensitivity (2)	[redacted]	NR

SAE	BLAST MT103-203 (n=116)		MT103-202 (Pilot) (n=21)	
Infections and infestations	[redacted]	Acinetobacter bacteraemia (1); atypical pneumonia (1); bacterial infection (1); bronchopneumonia (1); bronchopulmonary aspergillosis (1); cystitis klebsiella (1); device related infection (3); H1N1 (1); osteomyelitis (1); sepsis (1); sinusitis (2); staphylococcal infection (3); upper respiratory tract infection (1); urinary tract infection (1)	[redacted]	Bacterial sepsis (1); bronchopneumonia (1); catheter related infection (1); Escherichia sepsis (1)
Injury, poisoning and procedural complications	[redacted]	Accidental overdose (1); incision site haemorrhage (1); infusion related reaction (1); overdose (5); post lumbar puncture syndrome (1); spinal fracture (1); subdural haemorrhage (1); thermal burn (1)	[redacted]	Medical device complication (1); thrombosis in device (1)
Investigations	[redacted]	Alanine aminotransferase increased (2); aspartate aminotransferase increased (2); blood bilirubin increased (1); body temperature increased (1); c-reactive protein increased (4); hepatic enzyme increased (1); liver function test abnormal (1); prothrombin time prolonged (1)	[redacted]	NR
Neoplasms benign, malignant and unspecified (incl cysts and polyps)	[redacted]	Kaposi's sarcoma (1); leukaemia (1)	[redacted]	NR
Nervous system disorders	[redacted]	Aphasia (6); ataxia (2); cognitive disorder (1); dysarthria (1); encephalopathy (6); generalised tonic-clonic seizure (1); headache (2); intention tremor (1); leukoencephalopathy (1); motor dysfunction (1); paraesthesia (1); seizure (3); tremor (8)	[redacted]	Convulsion (1); epilepsy (1); somnolence (1); syncope (1)
Psychiatric disorders	[redacted]	Agitation (1); confusional state (1); disorientation (1)	[redacted]	NR
Skin and subcutaneous tissue disorders	[redacted]	Dermatitis contact (1); rash maculo-papular (1)	[redacted]	NR
Vascular disorders	[redacted]	Hypotension (1); thrombosis (1); vena cava thrombosis (1)	[redacted]	Hypertension (1)

NR - not reported

Table 15: Comparison of SAEs in MRD+BCP-ALL patients (Pooled BLAST + MT103-202) with known safety profile from relapsed and refractory Ph-BCP-ALL patients (Pooled MT103-206 + MT103-211 + TOWER)

Event	MRD+BCP-ALL patients (Pooled BLAST + MT103-202), n=137	Ph-BCP-ALL patients (Pooled MT103-206 + MT103-211 + TOWER), n=NR*
Treatment-emergent grade ≥3 AEs	[redacted]	[redacted]
Treatment-emergent serious AEs	[redacted]	[redacted]
Treatment-related AEs	[redacted]	[redacted]
Treatment-related serious AEs	[redacted]	[redacted]
Treatment-emergent EOIs grade ≥3	[redacted]	[redacted]
Treatment-emergent EOIs grade ≥4	[redacted]	[redacted]
EOI grade ≥3	[redacted]	[redacted]
EOI grade ≥4	[redacted]	[redacted]
Neurological AEs	[redacted]	[redacted]
Cytokine release syndrome	[redacted]	[redacted]
Medication errors	[redacted]	[redacted]

* *Total pooled n not reported. Individual studies reported as n=36 (MT103-206), n=189 (MT103-211), n=NR (TOWER)* [43]

4.3 Study included as comparator

Study 20120148 was a retrospective study that collected data on PH-BCP-ALL patients who were in complete haematological remission with MRD (see Table 16). The rationale for the study was to provide a frame of reference from which to compare the single-arm BLAST study of blinatumomab.[33] Treatment and outcome data were collected retrospectively from study groups across Europe and Russia (see CS,[1] Section B.2.9). MRD assessment was by PCR or by flow cytometry at a reference lab.[33] Study 20120148 collected OS and RFS data, but did not provide data on AEs. Within the limitations of the study design, the study was well conducted; as noted in Section 4.1.4, single-arm and retrospective studies are associated with known biases.

Eligibility criteria for Study 20120148 were available from the CS[1] (Section B.2.9) and the US NIH clinical trials registry;[41] these are presented below.

Study 20120148 inclusion criteria

Patients with Ph- BCP-ALL with haematological CR (defined as less than 5% blasts in bone marrow after at least 3 intensive chemotherapy blocks, and who met the following criteria:

- Detection of MRD (molecular failure or molecular relapse) at a level of $\geq 10^{-4}$ by PCR or $\geq 10^{-3}$ by flow cytometry at a reference lab

- Age 15+ at time of initial diagnosis of ALL. For patients 15-17 years of age at diagnosis, patients were not allowed to be enrolled in a paediatric trial, i.e. had to be treated according to adult protocols
- Initial diagnosis of ALL in the year 2000 or later
- History of ALL treatment (including response to first therapy, number of prior relapses) is available
- Relapse status and disease follow-up after time point of MRD detection is available.

Study 20120148 exclusion criteria

- Patients with extramedullary disease at timepoint of MRD detection
- Use of blinatumomab within 18 months of MRD detection
- Allogeneic HSCT prior to MRD detection at required level.

From the data collected in Study 20120148, a direct comparison analysis set (DCAS) was selected to act as matched controls for the BLAST study (see CS,[1] Section B.2.9). Additional criteria were applied in order to produce the DCAS. Data from Russian patients were excluded because MRD levels were not quantified. Patients were in their first haematological remission (CR1) only. Only patients aged 18 years or older at the MRD baseline date were included. Time to relapse had to be greater than 14 days from the date of MRD detection.

Baseline characteristics

Baseline characteristics for patients in Study 20120148 are presented in Table 17. Most of the patients in the DCAS were from [redacted] (see CS,[1] Section B.2.9.3). The DCAS included [redacted] from the UK.[33]

Table 16: Characteristics of retrospective control study

Study	Reference(s)	Study design	Population	Number of patients	Intervention	Date of initial diagnosis	Outcomes
Study 20120148 NCT02010931	Amgen Inc. Observational Research Study Report 2017[33]	Retrospective, international, multicentre	Adult Ph- BCP-ALL patients in haematological CR with MRD	Data collected for 287 patients [redacted] patients selected for DCAS	Standard care chemotherapy regimens, according to national treatment or study group protocols	2000 - 2014	Haematological RFS rate, OS, Mortality rate 100-days following HSCT

Information from CS Sections B.2.2 and B.2.9, CS clarification response[17] question A13 and Amgen Study Report[33]
DCAS= direct comparison analysis set

Table 17: Baseline characteristics of Study 20120148 direct comparison analysis set

Demographic	Study 20120148
	[redacted]* Prior to adjustment**
Male sex, n (%)	[redacted]
Median age (range), years	[redacted]
Age, n (%) ≥18 to <35 years	[redacted]
≥35 to <55 years	[redacted]
≥55 to <65 years	[redacted]
≥65 years	[redacted]
Relapse history, n (%) First CR	[redacted]
Second CR	[redacted]
Third CR	[redacted]
Baseline MRD levels, n (%) $\geq 10^{-1} < 1$	[redacted]
$\geq 10^{-2} < 10^{-1}$	[redacted]
$\geq 10^{-3} < 10^{-2}$	[redacted]
$< 10^{-3}$	[redacted]
Philadelphia chromosome disease status Negative	[redacted]
Confirmed t(4;11) Translocation / MLL-AF4+	[redacted]
Time from diagnosis to baseline (months) mean (SD)***	[redacted]
WBC at diagnosis ($\geq$30,000/mm^3)	[redacted]

Adapted from CS Section B.2.9 Table 28 and Appendix L Table 86 and Amgen Study report [44]
**Patients ≥18 years old with MRD load ≥1 × 10−3 detected by FC or PCR in CR1, time to haematological relapse >14 days after MRD diagnosis. **For details on adjustment see ERG report Section 4.4.*
****Time from initial diagnosis to baseline MRD status defined as the earliest MRD detection date following complete remission after at least three blocks of chemotherapy[33]*
CR: complete remission; DCAS: direct comparison analysis set. WBC: white blood cell

4.4 Indirect comparison

Owing to the lack of randomised data to inform the comparative effectiveness of blinatumomab versus standard care chemotherapy, the company performed an analysis based on the historical cohort DCAS, designed *post hoc* to include patients resembling those enrolled into the BLAST study. RFS and OS outcomes were considered; other outcomes listed in the final NICE scope[22] were not reported for the indirect comparison. Propensity score methods based on inverse probability of treatment weighting (IPTW) were used.

The data used to inform the analysis are described in Section 4.4.1. The methods used to estimate treatment effectiveness are described in Section 4.4.2 and are subsequently critiqued according to the items in the Quality of Effectiveness Estimates from Non-randomised Studies (QuEENS) checklist.[45]

4.4.1 Critique of included studies

The effectiveness of blinatumomab was informed by the BLAST study (n=[redacted]), as summarised in Section 4.2, whilst the effectiveness of standard care was informed by the historical comparator DCAS (n=[redacted]), as summarised in Section 4.3. Baseline characteristics of the full BLAST study and historical control DCAS are compared in Table 28 of the CS. As noted by the company, there were key differences between the two populations in terms of Ph status and relapse history.[1]

The BLAST primary analysis set (PAS) was trimmed to overlap with the historical comparator DCAS. The two key criteria defining this subgroup are the restriction to CR1 and Ph- individuals only; the full criteria are listed below:

- Ph- BCP- ALL;
- First complete haematological remission (CR1);
- MRD+ at a level of >1 x 10-3;
- ≥18 years old at MRD positivity (historical control study [Study 20120148]) or first blinatumomab treatment (BLAST [Study MT103-203]);
- Complete baseline covariate set;
- Time to relapse greater than 14 days from MRD detection (applied to historical control study data).

Trimming resulted in a subgroup of [redacted] patients for the BLAST PAS.

The timing of MRD assessment following diagnosis also varied between the BLAST PAS and historical comparator DCAS, and within different study groups contributing to the historical comparator DCAS. In order to align the baseline dates and to reduce bias due to the definition of MRD baseline date, patients in the historical comparator DCAS were excluded if their time to relapse was less than 14 days (the median time between MRD detection and first blinatumomab dose for BLAST patients). The baseline date for patients within the historical comparator study was set equal to their MRD detection date plus 14 days. This led to the exclusion of four patients from the historical control study, due to relapse during the first 14 days after MRD baseline (see company's clarification response,[17] question A11).

The cases from the control study were recruited from the year 2000 onwards, as opposed to the BLAST study, in which cases were recruited from 2010 onwards. There have been some changes to induction treatment, which may mean that more recently treated patients have lower rates of MRD positivity and lower rates of relapse. However, there is an absence of evidence for this in UK-treated patients. Clinical advice received by the ERG suggests that it is broadly reasonable to assume that treatments received by patients from 2000 onwards would be similar to current practice.

4.4.2. Critique of methods for estimating comparative effectiveness

Description of analysis performed by company

Differences between the BLAST PAS and historical comparator DCAS with respect to key baseline characteristics (prior to propensity score adjustment) are shown in Table 18. Balance with respect to individual covariates was assessed by the company using two methods: (i) univariate regression models were constructed to investigate the association between the treatment group (as the predictor), on each baseline characteristic (as the outcome variable) individually, using linear and logistic regression for continuous and binary baseline characteristics respectively, with results reported as *p*-values, and (ii) standardised mean differences between the two groups were calculated (formulae presented in CS Appendix L). The CS states that the criteria for concluding that adequate balance was achieved were: (i) non-significant *p*-values and (ii) standardised differences less than 0.2, with "best balance" achieved with standardised differences less than 0.1.[1]

Before applying the propensity score weighting, four of the listed covariates had *p*-values which were less than 0.05: age; country; time from diagnosis to baseline (months), and prior chemotherapy. The absolute standardised differences ranged from 0 to 0.56, with standardised differences greater than 0.2 observed for WBC at diagnosis (continuous) in addition to the four covariates listed above. Only two covariates (gender and T411mll4 mutation) exhibited standardised difference less than 0.1, which is indicative of good balance between the groups.

Due to the observed differences in baseline characteristics between the BLAST PAS and the historical control DCAS, IPTW based on a propensity score model was used. The overall aim of the procedure is to create balance between the two groups by producing a weighted sample that mimics the effect of randomisation in an RCT.[45] The propensity score model estimates the probability of being assigned to the treatment group as a function of a set of observable covariates. These propensity scores are used to construct weights that are applied to the observed data. Several weighting schemes may be considered, each of which results in different interpretations of the resulting treatment effect. The average treatment effect (ATE) measures the expected gain from the treatment for a randomly selected individual (across both samples) and is most appropriate when the treatment is relevant to the entire population represented by the data. Weights are applied to both the BLAST PAS and the historical control DCAS patients (see CS,[1] Appendix L). The average treatment effect on the treated (ATT) is relevant when the interest lies on the effect of treatment only for those who are treated (rather than the population of both treated and untreated patients). No weighting is applied to the blinatumomab patients, whilst patients in the historical control arm are weighted to match those in the treated study.

The company used ATT weights to inform the health economic model (see Section 5). The justification for this was that results based on ATT weights can be generalised to the population of patients from BLAST, which represents the prospectively selected anticipated licensed population, rather than the combined population of the BLAST study and the historical control study. Results using ATE weights are presented in CS Appendix L, and were used for a sensitivity analysis. In order to adjust for potential instability caused by very large weights, stabilised weights (applied to both the ATT and ATE analyses) were presented by the company, whereby the weight is multiplied by the marginal probability of receiving the actual treatment received.[46] This results in a smaller effective sample size of [redacted] (see company's clarification response,[17] A8 Additional Query). The stabilised weights were used to produce the estimates of treatment effect presented in the clinical effectiveness section of the CS. The company acknowledged that there was a lack of consistency between the results presented in the clinical effectiveness section and those used to inform the health economic model, but stated that they are confident that the application of the stabilised ATT (sATT) weights to the cost-effectiveness analyses would have "*no impact*" (see company's clarification response,[17] question A11).

Candidate variables for the company's propensity score model were chosen through discussion amongst the study team and clinicians. As stated in the company's response to clarification[17] (question A8), the majority of covariates were chosen based on prognostic factors that have been identified for ALL in published literature and to account for potential regional differences in treatment practices. Candidate variables included: age at primary diagnosis; sex; country; presence and type of an cytogenetic and molecular aberrations; time from primary diagnosis to MRD baseline data (months); baseline MRD level (ordinal variable, treated as continuous in the model); WBCs at diagnosis, and type of prior chemotherapy (binary: GMALL, other). The final propensity score model was chosen by including all candidate variables and two-way interactions into a logistic regression with treatment as the binary response. A stepwise selection algorithm was used with inclusion into the final model based on statistical significance (p<0.30).

Table 18: Covariate balance between BLAST PAS and historical control study, before and after adjustment using ATT weights (reproduced from company's clarification response question A8)

Characteristic	Unweighted				IPTW			
Mean (SD)/n (%)	Control [redacted]	Blinatumomab [redacted]	Standard Difference	*p*-value	Control [redacted]	Blinatumomab [redacted]	Standard Difference	*p*-value
Age at primary diagnosis (years)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Gender (female)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Country (not Germany)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
MRD at Baseline (recoded)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Time from diagnosis to baseline (months)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
WBC at diagnosis (>30,000/mm^3)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
WBC at diagnosis (continuous, log10)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
T411mll4 mutation (Yes)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Prior chemotherapy (GMALL)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Table 19: Summary of propensity score model covariates (modified from company's propensity score analysis report)

Covariate	estimate (SE)	*p*-value
age at primary diagnosis (years)	[redacted]	[redacted]
time from diagnosis to baseline (months)	[redacted]	[redacted]
MRD level at baseline	[redacted]	[redacted]
type of prior chemotherapy	[redacted]	[redacted]
Not GMALL	[redacted]	[redacted]
time from diagnosis to baseline (months) x type of prior chemotherapy	[redacted]	[redacted]
Not GMALL	[redacted]	[redacted]

SE - standard error
p-value from Wald Chi-Square statistic

Balance diagnostics after applying ATT weights to the historical control DCAS are presented in Table 18, based on the company's clarification response[17] (question A8). After applying ATT IPTW weights, [redacted] Balance diagnostics for the sATT weights used to estimate the treatment effects are shown in the CS[1] (Table 86, page 220). [redacted]

The derived sATT propensity score weights were used to perform a weighted Cox proportional hazards analysis and therefore estimate the hazard ratio (HR), providing a treatment effect comparing blinatumomab to standard care. Analyses were conducted separately for both RFS and OS. The primary analysis considered just one covariate (allowing a treatment effect), and an additional analysis was conducted including a time-dependent covariate for HSCT to account for differences between transplant rates observed between BLAST and the historical cohort. Analyses were conducted in SAS. The adjusted Kaplan-Meier plots using ATT weights presented by the company are shown in Figure 4 and Figure 5 for RFS and OS, respectively; estimated treatment effects are summarised in Table 20.

[redacted]

[redacted] After applying sATT weights, the 18-month OS probability with standard care chemotherapy, without censoring for HSCT, was [redacted] (CS Appendix L) and the median OS was slightly longer than prior to weighting, at [redacted]. For the BLAST PAS with sATT weights [redacted] without censoring for HSCT, the 18-month OS probability was [redacted]). Median OS for the BLAST patients [redacted].

[redacted] After applying sATT weights, the 18-month RFS probability with standard care chemotherapy, without censoring for HSCT, was [redacted] (CS Appendix L) and the median RFS was [redacted] [redacted]). For the BLAST PAS [redacted] without censoring for HSCT, the [redacted] RFS probability was [redacted]). Median RFS for the BLAST patients was [redacted].

Table 20: Estimated treatment effects based on sATT weights

Outcome	Median (months)		HR (95% CI)	
	Standard care [redacted]	Blinatumomab [redacted]	primary analysis	covariate for HSCT
RFS	[redacted]	[redacted]	[redacted]	[redacted]
OS	[redacted]	[redacted]	[redacted]	[redacted]

[redacted] 4 [redacted]

[redacted]

Note that after application of the sATT weights, the effective sample sizes are [redacted] 5 [redacted]

[redacted]

Note that after application of the sATT weights, the effective sample sizes are [redacted]

Critique of the analysis

The ERG considers that the IPTW method used by the company is appropriate and that other methods are not suitable in this case due to the limited sample size. The method makes two important assumptions. Firstly, the methods assume that there is *no unobserved confounding* (also described as selection on observables). When estimating treatment effects based on non-randomised data it is possible that a patient received a particular treatment because of some (observable or unobservable) factors. Unless properly accounted for, this will lead to selection bias in the estimated treatment effect.[45] Selection on observables implies that all factors which determine treatment and are correlated with the outcome are observable, and hence can be accounted for in the propensity score model. There may be unobservable factors which determine treatment allocation, but these are not correlated with the outcome. Secondly, the *overlap assumption* is also required. This means that, for any combination of covariates, it is possible for individuals to be allocated to either the treatment or control group, ruling out the possibility that individuals with certain observable characteristics are always in one group and never in the other.[45] Weaker versions of both of these assumptions are required for the validity of the ATT weights, compared with the requirements of the ATE weights.

The analysis was based on a subset of individuals, the BLAST PAS, rather than the whole study population. This "trimming" is generally required in order to meet the overlap assumption when the initial overlap between the two populations is poor.[45] However, this redefines the interpretation of the estimated treatment effects. The ATT weights presented as the company's primary analysis represent the average treatment effect for the population of the BLAST PAS (n=[redacted]), which was chosen to overlap with the historical control study, rather than the full BLAST study population. The company's justification of the choice of ATT weights (rather that the ATE weights that were pre-specified in the protocol) due to the BLAST study being in line with the anticipated marketing authorisation is therefore not consistent with the interpretation of the resulting estimates, which are representative of the subpopulation only.

The assumptions required by each of the weighting methods are described in the CS, however, it is not clearly stated whether there is reason to believe that the stronger assumptions required for the validity of the ATE weights may not be met. ATT weights were used in the company's health economic base case analysis despite the fact that there was "*less improvement in covariate balance after weighting when using ATT*".[44] Overall, the ERG does not consider that the company's choice of weights for the base case analysis has been clearly justified. There was also a lack of clarity and consistency caused by the use of sATT weights to estimate treatment effects, and the application of standard ATT weights in the cost-effectiveness analysis. The ERG considers that that the use of the standard ATT weights was appropriate.

Clinical advisors to the ERG considered that the candidate variables considered by the company were generally appropriate; however, they drew attention to the potential for unobserved confounders related to HSCT status. As noted within the CS[1] (Section B.2.9.5, page 75), transplanted patients may be systematically different in terms of both measured and unmeasured factors (such as availability of a suitable donor). The HSCT rate is higher in the BLAST study ([redacted]) than the historical control study ([redacted]), and the CS states that the comparison is vulnerable to HSCT being a confounding factor.

The ERG believes that the choice of a logistic regression model was appropriate. However, the inclusion of covariates in the final model was based on statistical significance only. The CS does not present any checks (e.g. model diagnostic plots) for the final model. After applying ATT weights to the historical control DCAS, the company's pre-specified criteria for judging balance between the populations was met. This was not true for the sATT weights used to estimate treatment effects, as three covariates (age, time from diagnosis to baseline, WBC at diagnosis) still had standardised differences greater than 0.2. However these results were not used to inform the cost-effectiveness analysis.

Furthermore, it should be emphasised that the propensity score weights (hence, also the adjusted Kaplan-Meier survival curves) are estimates with associated measures of uncertainty (e.g. SEs). It is unclear (although unlikely) that this has been accounted for in the estimation of treatment effects, hence the reported confidence intervals of the treatment effects are likely to underestimate the associated uncertainty. The ERG therefore considers that the reported treatment effects should be interpreted with caution.

4.5 Additional work on clinical effectiveness undertaken by the ERG

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

4.6 Conclusions of the clinical effectiveness section

4.6.1 Completeness of the CS with regard to relevant clinical studies and relevant data within those studies

The clinical evidence presented in the CS is based on a systematic review of adult BCP-ALL patients with MRD positivity after treatment. The company's study selection eligibility criteria were consistent with the decision problem outlined in the final NICE scope,[22] except that the comparator "monitor for relapse" was not included; the ERG's clinical advisors noted that some older and less fit patients may not be able to receive HSCT or to tolerate chemotherapy, but may be able to tolerate blinatumomab. It is unclear whether any relevant comparator data exist within this subgroup. Overall, the ERG believes that whilst the searches conducted by the company were flawed, it is unlikely that any relevant studies of blinatumomab in adult BCP-ALL patients with MRD positivity after treatment have been missed.

Evidence of the effectiveness of blinatumomab was provided from two single-arm open-label studies, BLAST (n=116) and MT103-202 (n=20), with no internal control group against which to estimate a treatment effect. Comparator data relating to standard care chemotherapy were provided from one historical control study, Study 20120148 (n=287), that analysed data from existing clinical databases.

AE data for blinatumomab were presented for BLAST and MT103-202. There were no data on AEs or HRQoL from historical control study 20120148.

4.6.2 Summary of clinical effectiveness outcomes reported in the CS in relation to relevant population, interventions, comparator and outcomes

Clinical advice received by the ERG suggested that baseline demographics and prior treatment in the BLAST study were broadly generalisable to the population of MRD+BCP-ALL patients in England.

The ERG notes that there will be a small population of patients who are unable to undergo HSCT or to tolerate chemotherapy who are unlikely to be represented within the BLAST population.

From the 116 patients in BLAST, median OS was [redacted], with an OS at 18 months follow-up of [redacted]. From 110 patients providing RFS data from BLAST, median RFS was [redacted]; RFS at 18 months was [redacted]. BLAST measured HRQoL using the EORTC QLQ-C30 and the EQ-5D. Based on the EORTC QLQ-C30, patients reported [redacted]. By the end of the core study, [redacted]

HRQoL as measured by EQ-5D did not change significantly from baseline to the end of the core study. [redacted] participants experienced at least one treatment-emergent AE. Events occurring in ≥20% of participants included: [redacted]. The most common EOIs of blinatumomab were: neurological events [redacted].

Comparative effectiveness was estimated by applying sATT propensity score weights to the standard care chemotherapy arm. Due to differences between the populations of BLAST and the historical control study, this was based on a subset of the original study populations which were restricted to Ph- and CR1 individuals only (BLAST PAS n=[redacted], historical control DCAS n=[redacted]). The resulting treatment effect estimates therefore reflect a narrower population than that defined in the final NICE scope[22] and the wording of the anticipated marketing authorisation.[24]

For the BLAST PAS, the 18-month OS probability was [redacted].

For the BLAST PAS, the 18-month RFS probability was [redacted]

4.6.3 Uncertainties surrounding the reliability of the clinical effectiveness evidence

A key limitation of the effectiveness evidence is the design of the included studies. The two blinatumomab studies were well conducted, however single-arm studies are subject to several biases.[34] Comparative effectiveness was estimated using propensity score methods which the ERG considers to have been appropriately applied by the company; however, the estimation of treatment effects based on non-randomised data is still subject to limitations, namely that it is not possible to account for unobserved confounders, and the company states that the comparison is vulnerable to HSCT being a confounding factor.

Treatment effects (HR) appear to have been calculated ignoring the uncertainty associated with the estimated propensity score weights, and therefore it is likely that the estimates presented within the CS underestimate the total uncertainty of the reported HR, resulting in erroneously narrow confidence intervals. The ERG therefore considers that the reported treatment effects should be interpreted with caution.

A further limitation of the available evidence relates to generalisability to the full population outlined in the final NICE scope and the anticipated license.[22] On the basis of clinical advice, the ERG considers the population characteristics of the BLAST PAS and the historical control DCAS to be representative of Ph- CR1 patients with MRD+ BCP-ALL. However, there is no evidence to inform the comparative effectiveness of blinatumomab compared with standard care chemotherapy in patients with CR2+ and/or Ph+ disease. In addition, no evidence is reported for blinatumomab versus monitoring for patients who are unable to receive HSCT or to tolerate chemotherapy but who would be able to tolerate blinatumomab.

5. COST EFFECTIVENESS

This chapter presents a summary and critical appraisal of the methods and results of the company's review of published economic evaluations and the *de novo* health economic analysis presented within the CS.[1] All analyses presented in this chapter including the Patient Access Scheme for blinatumomab.

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

5.1.1 Description of company's systematic review of cost-effectiveness evidence

A single search strategy reported in Appendix C (also used in the identification, selection and synthesis of clinical evidence) of the CS was used to identify the following study types: (i) economic analyses of all interventional therapies for adult ALL patients with MRD; (ii) HRQoL studies in patients with MRD+ BCP-ALL, and (iii) studies assessing the economic burden of patients with MRD+ BCP-ALL. The search strategies in both the database and website searches were fully reported. The records retrieved from the search were for all MRD+ ALL patients.

The following sources were searched: MEDLINE in Process [via PubMed], EMBASE [host not reported], Cochrane Database of Systematic Review [via Wiley Online Library], Cochrane Central Register of Controlled Trials [via Wiley Online Library], Database of Abstracts of Reviews of Effectiveness [via CRD], NHS Economic Evaluation Database [via CRD] and the Health Technology Assessment database [via CRD].

The ERG's concerns regarding the limitations of the company restrictions applied to the search strategy (MRD terms, study design and language limits) have been previously described in Section 4.1.1. Following the consultation with clinical experts, the ERG considers that the search is sufficiently comprehensive to retrieve important citations relating to all eligible studies.

The company's inclusion and exclusion criteria are reproduced in Table 21. The company's review included adult ALL patients with MRD-positivity after treatment and was not restricted by intervention. However, the company's searches did not identify any existing economic evidence relating to adult ALL patients with MRD-positivity after treatment.

Table 21: Company's review of existing economic studies - inclusion and exclusion criteria (adapted from CS, Appendix G)

	Inclusion criteria	Exclusion criteria
Population	Adult ALL patients with MRD-positivity after treatment	Paediatric patients MRD- ALL patients
Intervention/ comparator	Any interventional therapies	None
Outcomes	Cost effectiveness Measures of cost effectiveness (e.g. cost per QALY gained)	Non-economic outcomes
Study design	Economic analyses and HTA reports	Non-economic study designs

5.2 Description of the company's model

5.2.1 Model scope

As part of its submission to NICE, the company submitted a fully executable health economic model programmed in Microsoft Excel®. The scope of the company's health economic analysis is summarised in Table 22. The company's model assesses the cost-effectiveness of blinatumomab versus standard care chemotherapy in adult patients with Ph- MRD+ BCP-ALL in CR1. The incremental health gains, costs and cost-effectiveness of blinatumomab versus standard care are evaluated over a 50-year time horizon from the perspective of the UK NHS and Personal Social Services (PSS). Cost-effectiveness is expressed in terms of the incremental cost per quality-adjusted life year (QALY) gained. All costs and health outcomes are discounted at a rate of 3.5% per annum. Unit costs are valued at 2015/16 prices.

Table 22: Summary of company's health economic model scope

Population	Patients with Ph- MRD+ BCP-ALL in CR1
Intervention	Blinatumomab (up to 4 cycles)*
Comparator	Standard care - chemotherapy regimen assumed be comprised of vincristine, prednisolone, mercaptopurine, methotrexate and prophylaxis against CNS relapse using intrathecal methotrexate (treatment up to 2 years)
Primary health economic outcome	Incremental cost per QALY gained
Perspective	NHS and PSS
Time horizon	50 years
Discount rate	3.5% per annum
Price year	2015/2016

** All patients receiving blinatumomab are also assumed to receive prophylaxis against CNS relapse*
NHS – National Health Service; PSS – Personal Social Services

Population

The population considered within the company's health economic model is defined according to the characteristics of patients enrolled into the BLAST study and the historical comparator study who met the criteria stated in Section 4.4.1. These subgroups of the full study populations are described as the historical comparator DCAS ([redacted]) and the BLAST PAS ([redacted]). The company's health economic analysis is based on ATT propensity weights, rather than the ATE weights that are presented in CS

Appendix L. This approach was taken on the basis that the analysis based on the ATT weights *"can be generalised to the population of patients in BLAST rather than the combined populations of the BLAST and historical control studies"* (see Section 4.4).

It should be noted that the company's health economic analysis reflects a population of patients who are likely to be able to tolerate chemotherapy; clinical advisors to the ERG noted that owing to its toxicity profile, blinatumomab may be a treatment option for patients who are not fit enough to receive HSCT or to tolerate cytotoxic therapy; this subgroup is unlikely to be reflected by the population captured within the company's model. In addition, the company's economic analysis excludes two further subgroups of patients who were included in the BLAST study: (i) patients with Ph- MRD+ BCP-ALL who are in second or subsequent haematological remission (CR2+), and (ii) patients with Ph+ MRD+ BCP-ALL. The population considered within the model is therefore narrower than the anticipated marketing authorisation for blinatumomab (treatment of adults with MRD+ BCP-ALL).[24] The CS states that undertaking a formal economic analysis of blinatumomab in the broader patient population, which also includes patients in CR2+, was infeasible due to a lack of comparator data. However, despite the absence of any clinical or economic evidence to support the analysis of blinatumomab in these missing subgroups, the CS states *"due to the substantial unmet need across all sub-populations blinatumomab should be considered for use in alignment with its full anticipated marketing authorisation"* (CS,[1] page 15). The CS also states that it anticipates that blinatumomab would be used as early as possible in the treatment pathway. These issues are discussed further in Section 5.3.

Intervention

In the BLAST study, blinatumomab was administered as a continuous IV infusion at a dose of 15µg/m^2 per day for 4 weeks, followed by a 2-week treatment-free period. Patients could receive up to four consecutive treatment cycles of blinatumomab. In contrast, the model assumes that a single cycle of blinatumomab treatment is comprised of a continuous IV infusion at a dose of 28µg/day for 28 days, followed by a 14-day treatment-free interval. This is in line with the anticipated marketing authorisation for blinatumomab.[24] The model assumes that patients receive 1 cycle of induction treatment followed by up to 3 additional cycles of consolidation treatment.

Comparator

The comparator included in the company's model is standard care chemotherapy. Health outcomes for the comparator group are based on the historical control DCAS, whilst the costs of standard care are modelled according to the maintenance chemotherapy regimen for non-transplant patients used in the UKALL14 trial:[19]

- Vincristine 1.4mg/m^2 (maximum 2mg/dose) IV every 3 months for up to 2 years

- Prednisolone 60mg/m^2 orally 5 days every 3 months for up to 2 years
- Mercaptopurine 75mg/m^2 orally daily for up to 2 years
- Methotrexate 20mg/m^2 orally once weekly for up to 2 years
- Prophylaxis against CNS relapse using intrathecal methotrexate 12.5mg every 3 months for up to 2 years.[1]

The final NICE scope[22] also included a further comparator of "monitor for relapse" (no active treatment); the CS[1] justifies the exclusion of this comparator by stating: "*Based on expert clinical opinion it is highly unlikely that MRD+ patients who have a high risk of relapse would solely be monitored for relapse without any treatment*" (CS,[1] page 15). However, clinical advisors to the ERG noted that this comparator is relevant for those patients who are unable to undergo HSCT or to tolerate chemotherapy, but are able to tolerate blinatumomab. This comparator therefore should have been explored in the company's economic analysis.

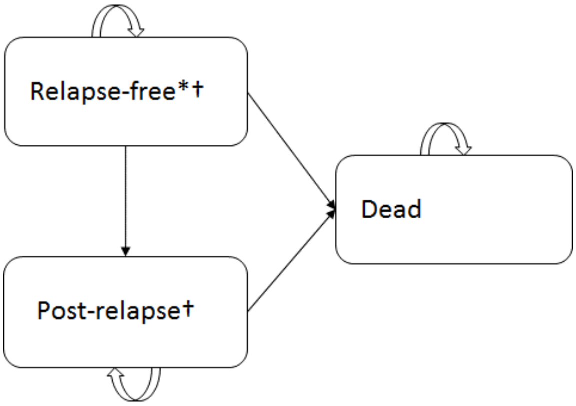
5.2.2 Description of the company's health economic model structure and logic

The company's model is comprised of a main structure which reflects RFS and OS outcomes from the BLAST PAS and historical control study DCAS, as well as two linked sub-models which estimate additional costs and HRQoL decrements associated with HSCT given before and/or after relapse. The subsequent sections describe the main model structure and the two HSCT sub-models separately.

Main partitioned survival model structure

The company's model adopts a partitioned survival approach based on three health states: (1) relapse-free; (2) post-relapse, and (3) dead (see Figure 6). Patients enter the model in the relapse-free state with an initial age of 45.38 years. Health state transitions are estimated over a total of 2,607 weekly cycles (approximately 50 years); at this timepoint, more than 99.9% of patients in each treatment group have died. The probability of being alive and relapse-free at any time t is based on a parametric (Gompertz) model fitted to the treatment-specific RFS time-to-event data from the BLAST PAS and the historical control DCAS with ATT weights. The probability of being alive at any time t is modelled using a parametric (log normal) mixture cure model fitted to the OS time-to-event data from the BLAST PAS and the historical control DCAS with ATT weights, as well as a separately estimated general population survivor function. The latter OS survivor function is estimated using age- and sex-specific mortality risks from life tables which are uplifted by a factor of 4 (based on the NICE Appraisal Committee's preferred assumption within the appraisal of inotuzumab ozogamicin for treating R/R ALL[47]) to reflect the potential long-term effects of complications of cytotoxic chemotherapy, and/or allogeneic HSCT on survival. Within the model trace, the probability of surviving during each model cycle is determined by the cumulative OS probability at the end of the previous model cycle and the maximum OS hazard for the current cycle derived from the parametric OS cure model and the uplifted general population survival curve. The probability of being alive and in the post-relapse state at any time *t* is calculated as the difference between the cumulative survival probabilities for OS and RFS.

Figure 6: Company's model structure



** RFS time divided into time on treatment and post-discontinuation*
† Patients may enter state-specific HSCT sub-model

Pre-relapse and post-relapse HSCT sub-models

Given the use of a partitioned survival approach in which health states are defined according to patients' survival and relapse status, the company's model structure does not explicitly account for differential RFS and OS impacts for those patients who receive HSCT; within the model, the proportionate use of pre-relapse HSCT is causally unrelated to RFS and OS events, whilst post-relapse HSCT use is partially dependent on RFS. In both treatment groups, the probability that a patient undergoes HSCT is approximated using separate pre-relapse and post-relapse HSCT sub-models in order to attribute costs and QALY losses associated with this intervention.

For patients who are relapse-free, the modelled (time-invariant) 6-monthly probability of receiving HSCT was calibrated such that the predicted cumulative probability of having undergone pre-relapse HSCT at 48 months matches the observed probability from the BLAST study and the historical control study. Beyond 48 months (based on the time of the last observed pre-relapse HSCT in BLAST and the historical control study), the model assumes that patients in the relapse-free health state of the main partitioned survival model cannot subsequently undergo HSCT, unless they relapse and enter into the post-relapse HSCT sub-model. Whilst the modelled proportion of patients receiving pre-relapse HSCT is dependent on the RFS function, OS in the main partitioned survival model is unaffected by the pre-relapse HSCT sub-model. After undergoing HSCT, 6-monthly follow-up costs and QALY losses are estimated using HSCT follow-up data from NHS Blood and Transplant[48] up to 2-years, and using uplifted general population survival rates thereafter.

With respect to the post-relapse HSCT sub-model, the per cycle probability of receiving HSCT after relapse is calculated within the model by determining the number of RFS events since the end of the previous 6-month HSCT sub-model cycle (derived from the modelled RFS curve) and a time-invariant treatment group-specific probability that an RFS event is death. The 6-month probability of undergoing post-relapse HSCT is determined by two factors: (i) the probability of undergoing HSCT for those patients who have not previously undergone HSCT whilst relapse-free, and (ii) the probability of undergoing HSCT for those patients who have previously undergone HSCT whilst relapse-free. As the model structure does not capture a patient's history of HSCT in the pre-relapse state, the model necessarily employs an assumption which attempts to estimate the probability of receiving post-relapse HSCT according to whether patients have undergone pre-relapse HSCT or not. In simple terms, the model is intended to assume that patients with pre-relapse HSCT do not relapse until all patients without pre-relapse HSCT have relapsed (see company's clarification response,[17] question B32, although the ERG notes that the implementation actually requires further assumptions about when the HSCT probability switches). As with the pre-relapse HSCT sub-model, 6-monthly follow-up costs and QALY losses are estimated using HSCT follow-up data from NHS Blood and Transplant[48] up to 2-years, and using uplifted general population survival rates thereafter.

Modelling HRQoL impacts

The model assumes that HRQoL is principally determined by relapse status, time spent alive and relapse-free and treatment received. Within the blinatumomab group, the model applies different health utilities in the relapse-free state over time; HRQoL is also assumed to differ for patients who are still receiving treatment and for those who have discontinued blinatumomab. Within the standard care group, the model applies fixed utilities for the relapse-free and relapsed states for up to 5-years. Within both treatment groups, HRQoL in the relapse-free state beyond 5-years is assumed to reflect that of the age- and sex- adjusted general population, less a constant utility decrement of 0.02, which is assumed to reflect long-term impacts associated with radiotherapy, chemotherapy, and HSCT. In addition, further time-dependent QALY losses are applied for those patients undergoing HSCT for up to 5 years. A further QALY loss is also applied to account for patients' proximity to death.

Modelled treatment pathway and associated costs

The company's model includes the following cost components: (i) drug acquisition; (ii) drug administration; (iii) health state resource use; (iv) HSCT; (v) salvage therapy costs, and (vi) a cost associated with death.

Within the blinatumomab group, the model assumes the following treatment pathway:

- Patients receive up to four cycles of blinatumomab irrespective of relapse status (experiencing relapse does not trigger the discontinuation of blinatumomab). Each cycle is comprised of 28 days receiving 28μg blinatumomab followed by 14 days without treatment. The model calculates blinatumomab costs based on the unweighted mean proportion of those patients starting the cycle and those still on treatment at the end of the cycle.
- Prophylaxis against CNS relapse is given to all patients in the relapse-free health state for up to 2 years, unless they progress to pre-relapse HSCT or die. The prophylaxis regimen is comprised of 15mg methotrexate, 40mg cytarabine and 4mg dexamethasone given once every 13 weeks. All regimen components are assumed to be administered by intrathecal injection during a single outpatient appointment.
- Patients are assumed to be eligible to receive HSCT pre-relapse and/or post-relapse. The precise resource use assumptions relating to the HSCT procedure and the initial 2-year follow-up period are not clear from the CS[1] or the source material cited therein.[48, 49] From 2 years after HSCT, patients in post-HSCT follow-up receive 100mg/day cyclosporine indefinitely, but do not incur any further costs associated with visits to health care practitioners. The proportion of patients remaining in HSCT follow-up is assumed to decline over time according to the estimated proportion of patients surviving.
- All patients who relapse receive salvage chemotherapy using FLAG-IDA. This regimen is assumed to be comprised of: filgrastim 0.005mg/Kg (9 days treatment per cycle); fludarabine 30mg/m^2 (5 days treatment per cycle); cytarabine 2,000mg/m^2 (5 days treatment per cycle), and idarubicin 8mg/m^2 (3 days treatment per cycle). The model assumes that 16.8 inpatient days are required to administer this regimen per FLAG-IDA cycle (cycle duration not reported in the CS[1]). Thirty-seven percent of patients who receive one round of salvage chemotherapy are assumed to subsequently receive a further round of the same regimen.

Within the standard care group, the model assumes the following treatment pathway:

- All patients receive chemotherapy whilst relapse-free for up to 2 years unless they undergo pre-relapse HSCT (at which point, chemotherapy is assumed to be discontinued), relapse or die. This treatment is costed according to the maintenance regimen for the non-transplanted population of the UKALL14 trial.[19] This regimen is assumed to be comprised of: (i) vincristine (IV, 1.4mg/m^2 once every 13 weeks); (ii) methotrexate (intrathecal, 12.5mg once every 13 weeks); (iii) prednisolone (oral, 60mg/m^2 5 times every 13 weeks); (iv) mercaptopurine (oral, 75mg/m^2 daily) and (v) methotrexate (oral, 20mg/m^2 weekly).
- HSCT is modelled using the same approach as in the blinatumomab group.
- Salvage chemotherapy is modelled using the same approach as in the blinatumomab group.

The application of different RFS and OS time-to-event curves leads to different trajectories through the main model health states, which when combined with assumptions regarding HSCT use and associated health losses and costs, produce different profiles of total costs and health outcomes for the two treatment groups. Incremental cost-effectiveness is calculated in a pairwise fashion as the difference in costs divided by the difference in QALYs for blinatumomab and standard care.

5.2.3 Key structural assumptions employed within the company's model

The company's model employs the following structural assumptions:

- All patients enter the model in the relapse-free health state.
- HRQoL is principally determined by relapse status, sojourn time in the relapse-free state and treatment group (the latter is driven largely by the treatment-related MRD response rate).
- Blinatumomab is assumed to be continued for up to four six-weekly cycles. Adjunctive prophylaxis against CNS relapse is assumed to be continued for up to nine quarterly cycles, or until HSCT, incidence of relapse, or death.
- Standard care chemotherapy is assumed to be continued for up to eight quarterly cycles, or until HSCT, incidence of relapse, or death.
- The RFS hazard is assumed to follow a Gompertz distribution in both groups (using an approach which is analogous to fitting models independently to each treatment group).
- The OS hazard is assumed to follow a log normal mixture cure model in both groups (which allows a different cure fraction but has the same standard parametric model parameters between the treatment groups).
- The probability of undergoing pre-relapse HSCT is assumed to be constant with respect to time.
- If a patient does not relapse, they are assumed to only be eligible to receive HSCT within the first four years of entering the model.
- Prior to the point at which the proportion of patients who are relapse-free is less than or equal to the cumulative proportion of patients who received a HSCT pre-relapse, all patients who relapse are assumed to have not received a pre-relapse HSCT; after this point, all patients who relapse are assumed to have received a pre-relapse HSCT.

5.2.4 Evidence used to inform the company's model parameters

The main groups of model parameters and the evidence sources used to populate these are summarised in Table 23. These are discussed in further detail in the subsequent sections.

Table 23: Evidence sources used to inform company's model parameters

Parameter type	Parameter	Source(s)
Time-to-event parameters	RFS - blinatumomab	BLAST PAS subgroup[1]
	RFS - standard care	Historical control study DCAS with ATT weights[1]
	OS - blinatumomab	BLAST PAS subgroup[1]
	OS - standard care	Historical control study DCAS with ATT weights[1]
Probability RFS event is death	RFS death probability - blinatumomab	BLAST PAS subgroup[1]
	RFS death probability – standard care	Historical control study DCAS with ATT weights[1]
HRQoL	Health utility – relapse-free ≤5 years	GLM/GEE regression based on BLAST data[1]
	Health utility – relapse-free >5 years (excluding additional HRQoL decrement for cured population)	Kind *et al*[50]
	Health utility – relapsed	Logistic regression using matched patients from BLAST and TOWER subgroups[1]
	QALY loss - HSCT (time-dependent)	Kurosawa *et al*[51]
	QALY loss – proximity to death	GLM/GEE regression based on BLAST data[1]
	Utility decrement for cured population – exposure to radiotherapy, chemotherapy, and HSCT	Assumption based on BLAST GLM/GEE[1] and Kind *et al*[50]
Mean dosing	Proportion of patients receiving blinatumomab dose during each treatment cycle (up to 4 doses)	BLAST[1]*
Probability of receiving HSCT	Probability of receiving pre-relapse HSCT	Calibrated to 4-year data from BLAST (blinatumomab) and historical control (standard care)
	Probability of receiving post-relapse HSCT	Estimated using BLAST and Study NCT02003612 (same probabilities used in each group)
Resource use and costs	Inpatient and outpatient resource use for standard of care and patients discontinuing blinatumomab	Face-to-face interviews with clinical experts (n=2)[1]
	HSCT procedure and subsequent follow-up (0-24 months)	NHS Blood and Transplant.[48] Cyclosporine costs taken from the British National Formulary (BNF)[25]
	Maintenance chemotherapy (standard care group)	Based on subgroup of UKALL14.[19] Unit costs taken from eMIT[52]
	Salvage chemotherapy	NICE TA450[53] (blinatumomab for relapsed/refractory ALL)
	Terminal care costs	King's Fund and Marie Curie reports[54, 55]
	Prophylaxis against CNS relapse for patients receiving blinatumomab	eMIT[52]
	Blinatumomab acquisition cost (including PAS)	Amgen[1]
	Unit costs for visits, appointments, hospitalisations, laboratory tests, radiological tests and AEs	NHS Reference Costs 2015/16[56]

** assumes 50% drug costs for those discontinuing within each cycle*
PAS – primary analysis set; DCAS – direct comparison analysis set; ATT - average treatment effect on the treated; GLM/GEE – generalised linear model/generalised estimating equation; eMIT – Electronic Market Information Tool; TA – technology appraisal

Time-to-event analysis

The company fitted parametric survival curves to time-to-event data from the BLAST PAS and the ATT weighted historical control DCAS. RFS for patients in the blinatumomab group was defined as the interval from the date of first blinatumomab treatment for MT103-203 patients from BLAST until haematological relapse or death (whichever occurred first). In order to avoid an immortal time bias (whereby a patient experiences an event before they are at risk within the study), the RFS interval for the historical comparator patients was adjusted to exclude patients with a time to relapse of less than 14 days (the median time between MRD detection and first blinatumomab dose for BLAST patients); the baseline date for patients within the historical comparator study was set equal to their MRD detection date plus 14 days. OS outcomes for patients in BLAST and the historical control study also relate to these same baseline timepoints, but include only death as an event.

A large range of survival models were fitted to the available RFS and OS data, including: (i) standard parametric models, (ii) restricted cubic spline (RCS) models, and (iii) mixture/non-mixture cure models (see Table 24). For most of the model types considered, the company fitted joint models which include a treatment effect covariate (an HR or constant acceleration factor; referred to in the CS as "restricted" models) and independent models which include treatment group interaction terms for every distributional parameter and are thus equivalent to fitting separate models to the treatment and control groups (referred to in the CS as "unrestricted" models). In addition, the cure models include both unrestricted and restricted model forms as well as third model type which allows a different cure fraction (θ) for the two groups, but the standard model parameters are otherwise the same for the remaining uncured population. This "cure" model form therefore implies that treatment group affects the likelihood of achieving a cure only, whilst for patients who are not cured, the time-to-event distribution is the same for both the standard care and blinatumomab treatment groups. For the RCS models, three variations were considered according to whether splines were fitted to the log cumulative hazard, log cumulative odds, or the inverse normal survival distribution. These are referred to by the company as the RCS Weibull, the RCS log logistic and the RCS log normal, respectively. Although it was not clear from the CS, the code provided by the company following the clarification process[17] (question B4) suggests that all RCS models assume one knot (where an increasing number of knots indicates a more flexible model). Thirty-eight models were fitted to the available RFS data. The same model forms were fitted to the OS data, however three of these (the gamma mixture cure, the gamma mixture cure [unrestricted] and the gamma non-mixture cure [unrestricted]) failed to converge, hence 35 models were fitted to the OS data.

Table 24: Summary of parametric models fitted to RFS and OS data

Standard parametric models	Flexible parametric models	Cure models
Exponential	RCS log logistic (R)	Gamma mixture cure*
Generalised F (R)	RCS log logistic (U)	Gamma mixture cure (R)*
Generalised F (U)	RCS log normal (R)	Gamma mixture cure (U)
Generalised gamma (R)	RCS log normal (U)	Gamma non-mixture cure
Generalised gamma (U)	RCS Weibull (R)	Gamma non-mixture cure (R)
Gompertz (R)	RCS Weibull (U)	Gamma non-mixture cure (U)*
Gompertz (U)	Piecewise exponential	Log normal mixture cure
Log logistic (R)		Log normal mixture cure (R)
Log logistic (U)		Log normal mixture cure (U)
Log normal (R)		Log normal non-mixture cure
Log normal (U)		Log normal non-mixture cure (R)
Weibull (R)		Log normal non-mixture cure (U)
Weibull (U)		Weibull mixture cure
		Weibull mixture cure (R)
		Weibull mixture cure (U)
		Weibull non-mixture cure
		Weibull non-mixture cure (R)
		Weibull non-mixture cure (U)

** Model presented for RFS analysis only*
R – restricted; U – unrestricted; RCS – restricted cubic spline

According to the CS, model discrimination was undertaken based on the consideration of five factors: (i) internal consistency; (ii) goodness-of-fit statistics; (iii) visual fit; (iv) evidence relating to underlying treatment effect, and (v) consistency with external data. The CS does not provide any information regarding the use of clinical judgement to assess the clinical plausibility of the extrapolated portions of the actual fitted parametric curves or their associated hazard functions.

Internal consistency of the RFS and OS models related to two considerations. Firstly, in instances in which the OS model under consideration and the selected base case RFS curves cross (thereby presenting a logically inconsistency), the OS model was excluded from further consideration. Secondly, the CS states that OS models were preferred if the difference in expected post-relapse survival gain between treatment groups was *"relatively small,"*, although little detail is provided in the CS regarding how this judgement was made.

Goodness-of-fit of the RFS and OS models was assessed using the Bayesian Information Criterion (BIC). According to the CS,[1] this measure was selected because it *"penalises overly complex models and its use mitigates risk of overfitting statistical noise in the tails of the observed distributions"* (CS,[1] page 94). Akaike Information Criterion (AIC) statistics for the fitted models were not presented.

Evidence relating to the underlying treatment effect between groups was based on consideration of counterfactual Kaplan-Meier survival plots (whereby estimated treatment effects are applied to the baseline Kaplan-Meier function) and examination of Schoenfeld residuals.[57] Other diagnostic plots (e.g., log cumulative hazard plots or their equivalents) were not presented.

External validity was assessed through comparison of predicted model outcomes with adjusted data from a meta-analysis of studies assessing the association between MRD status and clinical outcomes including EFS and OS in adults with ALL (Berry *et al*[18]). The data from Berry *et al* were used *"to assess the external validity of the RFS and OS distributions used in the model as well as the magnitude of the increase in RFS and OS that would be expected given the effect of blinatumomab on MRD response"* (company's clarification response,[17] question B12).

Given the wide range of parametric models included in the model-fitting process, the company considered only the five best fitting RFS models, determined according to their BIC; all other RFS models were excluded at this point. Similarly, the company considered only the five best fitting OS models which did not produce a logical inconsistency when viewed alongside the selected deterministic base case RFS curve. The other criteria for model choice described above were therefore considered only for these five best-fitting RFS and OS models. The ERG notes a lack of clarity within the CS regarding the company's subjective judgements of "good", "moderate" and "poor" in relation to these other model selection criteria. The company's clarification response[17] (question B16) provides additional detail and describes a "good" fit as *"the two curves are virtually the same, with no systematic over or under estimation"*, and a "poor" fit as *"the two curves are substantially different with apparent systematic over or underestimation over some range of the curve."*

Table 25 presents the BIC statistics for the 38 fitted RFS models. Table 26 presents the BIC statistics for the 35 fitted OS models. The five best fitting (and in the case of OS, logically consistent) models taken forward for further consideration by the company are highlighted in bold in each table.

Table 25: BIC statistics – RFS models

Parametric model	Model class	BIC	Considered further in the CS?
Exponential	Standard	1321.743	No
Generalised F (R)	Standard	1230.616	No
Generalised F (U)	Standard	1244.548	No
Generalised gamma (R)	Standard	1229.286	No
Generalised gamma (U)	Standard	1240.161	No
Gompertz (R)	**Standard**	**1222.061**	**Yes**
Gompertz (U)	**Standard**	**1225.587**	**Yes**
Log logistic (R)	Standard	1228.876	No
Log logistic (U)	Standard	1234.358	No
Log normal (R)	**Standard**	**1227.202**	**Yes**
Log normal (U)	Standard	1232.716	No
Weibull (R)	Standard	1257.919	No
Weibull (U)	Standard	1260.687	No
RCS log logistic (R)	**Flexible parametric**	**1225.662**	**Yes**
RCS log logistic (U)	Flexible parametric	1236.037	No
RCS log normal (R)	Flexible parametric	1229.012	No
RCS log normal (U)	Flexible parametric	1239.741	No
RCS Weibull (R)	Flexible parametric	1230.052	No
RCS Weibull (U)	Flexible parametric	1236.607	No
Piecewise exponential	Flexible parametric	1265.064	No
Gamma mixture cure	Cure	1236.985	No
Gamma mixture cure (R)	Cure	1233.307	No
Gamma mixture cure (U)	Cure	1244.343	No
Gamma non-mixture cure	Cure	1231.214	No
Gamma non-mixture cure (R)	Cure	1233.447	No
Gamma non-mixture cure (U)	Cure	1244.415	No
Log normal mixture cure	Cure	1233.42	No
Log normal mixture cure (R)	Cure	1228.875	No
Log normal mixture cure (U)	Cure	1234.392	No
Log normal non-mixture cure	Cure	1227.46	No
Log normal non-mixture cure (R)	Cure	1229.115	No
Log normal non-mixture cure (U)	Cure	1234.655	No
Weibull mixture cure	Cure	1235.512	No
Weibull mixture cure (R)	Cure	1234.439	No
Weibull mixture cure (U)	Cure	1238.882	No
Weibull non-mixture cure	**Cure**	**1227.299**	**Yes**
Weibull non-mixture cure (R)	Cure	1230.72	No
Weibull non-mixture cure (U)	Cure	1235.785	No

BIC – Bayesian Information Criterion; R – restricted; U – unrestricted; RCS – restricted cubic spline

Table 26: BIC statistics – OS models

Parametric model	Model class	BIC	Considered further in the CS?
Exponential	Standard	1197.457	No
Generalised F (R)	Standard	1176.196	No
Generalised F (U)	Standard	1190.688	No
Generalised gamma (R)	Standard	1173.349	No
Generalised gamma (U)	Standard	1183.772	No
Gompertz (R)	Standard	1181.63	No
Gompertz (U)	Standard	1187.016	No
Log logistic (R)	Standard	1179.883	No
Log logistic (U)	Standard	1185.326	No
Log normal (R)	Standard	1173.671	No
Log normal (U)	Standard	1179.173	No
Weibull (R)	Standard	1197.723	No
Weibull (U)	Standard	1201.822	No
RCS log logistic (R)	Flexible parametric	1169.497	No
RCS log logistic (U)	Flexible parametric	1180.351	No
RCS log normal (R)	Flexible parametric	1171.037	No
RCS log normal (U)	Flexible parametric	1181.938	No
RCS Weibull (R)	Flexible parametric	1169.987	No
RCS Weibull (U)	Flexible parametric	1180.608	No
Piecewise exponential	Flexible parametric	1196.289	No
Gamma mixture cure	Cure	Failed to converge	No
Gamma mixture cure (R)	Cure	Failed to converge	No
Gamma mixture cure (U)	Cure	1194.837	No
Gamma non-mixture cure	Cure	1177.058	No
Gamma non-mixture cure (R)	Cure	1182.231	No
Gamma non-mixture cure (U)	Cure	Failed to converge	No
Log normal mixture cure	**Cure**	**1173.187**	**Yes**
Log normal mixture cure (R)	Cure	1177.834	No
Log normal mixture cure (U)	**Cure**	**1182.969**	**Yes**
Log normal non-mixture cure	Cure	1171.676	No
Log normal non-mixture cure (R)	Cure	1176.964	No
Log normal non-mixture cure (U)	**Cure**	**1182.057**	**Yes**
Weibull mixture cure	**Cure**	**1188.202**	**Yes**
Weibull mixture cure (R)	Cure	1193.661	No
Weibull mixture cure (U)	Cure	1197.174	No
Weibull non-mixture cure	**Cure**	**1183.034**	**Yes**
Weibull non-mixture cure (R)	Cure	1188.552	No
Weibull non-mixture cure (U)	Cure	1192.722	No

BIC – Bayesian Information Criterion; R – restricted; U – unrestricted; RCS – restricted cubic spline

For RFS, the five best fitting models were: (i) Gompertz restricted; (ii) Gompertz unrestricted; (iii) RCS log-logistic; (iv) log normal, and (v) Weibull non-mixture cure. Table 27 summarises the company's judgements regarding model selection for these five best-fitting RFS models. The unrestricted Gompertz was selected for use in the base case analysis *"due to its good statistical fit, visual fit and external validity"* (CS,[1] page 101).

Table 27: Summary of model selection criteria for 5 best-fitting RFS models (adapted from CS Table 37)

Distribution	Δ BIC	Cure fraction	Treatment effect	Visual fit	External validity	Company comments
Gompertz (R)	--	Blin: 48.5% SC: 16.0%*	Moderate	Moderate	Good	Counterfactual plots suggest proportional hazards may overestimate long-term benefit of blinatumomab.
Gompertz (U)	**3.53**	**Blin: 39.5% SC: 17.2%***	**--**	**Good**	**Good**	**Good visual fit, statistical fit, and external validity.**
RCS log logistic (R)	3.6	Blin: 0% SC: 0%	Good	Moderate	Poor	Proportional odds model. Underestimates benefit of blinatumomab relative to external data.
Log normal (R)	5.14	Blin: 0% SC: 0%	Good	Poor	Poor	Accelerated failure time model. Poor visual fit, underestimates benefit of blinatumomab relative to external data.
Weibull non-mixture (cure)	5.24	Blin: 47.8% SC: 15.8%	Moderate	Moderate	Good	Treatment effect† parameterised as a cure model, but also follows proportional hazards. Counterfactual plots suggest proportional hazards may overestimate long-term benefit of blinatumomab.

** Not parameterised as cure models*
† The ERG is unclear about the meaning of treatment effect in this context as the cure model uses the same standard model parameters but a different cure fraction between groups
BIC - Bayesian Information Criterion

For OS, the five best fitting models which do not intersect the deterministic base case unrestricted Gompertz RFS curves were: (i) log normal mixture cure; (ii) log normal non-mixture cure unrestricted; (iii) log normal mixture cure unrestricted; (iv) Weibull non-mixture cure, and (v) Weibull mixture cure. Table 28 summarises the company's judgements regarding model selection for these OS models. The log normal mixture cure was selected for inclusion in the base case analysis as it had a *"much better statistical fit than the other distributions considered"* (CS,[1] page 108).

Table 28: Summary of model selection criteria for 5 best-fitting logically consistent OS models (adapted from CS Table 40)

Distribution	Δ BIC	Cure fraction	Treatment effect*	Visual fit	External validity	Δ PRS (years)	Company comments
Log normal mixture (Cure)	**--**	**Blin: 45.3% SC: 21.3%**	**--**	**Good**	**Moderate**	**-0.70**	**Best-fitting distribution among those consistent with base-case RFS. Large difference in BIC versus next best-fitting distribution.**
Log normal non-mixture (Cure, U)	8.87	Blin: 45.3% SC: 19.3%	--	Good	Moderate	-0.69	Poor statistical fit.
Log normal mixture (Cure, U)	9.78	Blin: 46.6% SC: 21.0%	--	Good	Moderate	-0.65	Poor statistical fit.
Weibull non-mixture (Cure)	9.85	Blin: 42.8% SC: 23.8%	Good	Good	Moderate	-1.58	Poor statistical fit. Treatment effect counterfactual plots are supportive of proportional hazards. Large difference in PRS.
Weibull mixture (Cure)	15.02	Blin: 46.8% SC: 24.9%	--	Good	Moderate	-1.11	Poor statistical fit. Large difference in PRS.

** The ERG is unclear how this could be assessed for cure models and notes that the fields for the log normal mixture cure and Weibull non-mixture cure models are blank*
BIC - Bayesian Information Criterion; PRS – post-relapse survival

Figure 7 presents a comparison of the empirical RFS Kaplan-Meier curves and the unrestricted Gompertz RFS models. Figure 8 presents a comparison of empirical OS Kaplan-Meier curves and log normal mixture cure OS models.

Figure 7: Comparison of empirical RFS Kaplan-Meier curves and RFS unrestricted Gompertz models

[redacted]

Figure 8: Comparison of empirical OS Kaplan-Meier curves and OS log normal mixture cure models

[redacted]

Proportion of RFS events that are deaths

The probability that an RFS event is death is assumed to differ between the treatment groups, based on the BLAST PAS and the ATT-weighted historical control DCAS. As shown in Table 29, 47.1% and 8.5% of RFS events were estimated to be deaths in the blinatumomab and standard care groups, respectively. The CS notes that the higher rate of deaths for blinatumomab may reflect: (i) the more frequent use of HSCT in BLAST; (ii) a *"notable"* proportion of BLAST patients undergoing transplants from mismatched donors thereby leading to greater risks of infection and death, and (iii) potentially incomplete reporting of HSCT receipt in BLAST.

Table 29: Percentage of RFS events which were deaths (reproduced from CS Table 38)

RFS events	BLAST (blinatumomab)		Historical control (standard care)	
	n	**%**	**n**	**%**
Unweighted				
Death	16	47.1%	14	10.7%
Relapse	18	52.9%	117	89.3%
Total	34	100.0%	131	100.0%
ATT-IPTW				

Death	16	47.1%	10.4	8.5%
Relapse	18	52.9%	112	91.6%
Total	34	100.0%	122.3	100.0%
ATE-IPTW				
Death	13.8	40.2%	13	10.1%
Relapse	20.5	59.8%	115.6	90.0%
Total	34.3	100.0%	128.5	100.0%

ATT - average treatment effect on the treated; ATE - average treatment effect; IPTW - inverse probability of treatment weighting

Health-related quality of life

Utility values for the pre-relapse states were based on EQ-5D utility values for patients included in the BLAST PAS (n=[redacted]). The company fitted a generalised linear model/generalised estimating equations (GLM/GEE) regression model with EQ-5D utility as the dependent variable, and covariates for baseline utility, a patient-level indicator variable of MRD response during cycles 1 or 2, a time-dependent indicator variable for on versus off treatment, and a time-dependent indicator variable for death within 6 months.[1] Patients without baseline assessments or any follow-up assessments were excluded from the model. In addition, utility assessments on or after relapse were also excluded from the analysis. A total of 63 patients from the BLAST PAS contributed data to the GLM/GEE model.

The CS states that post-relapse utility assessments in BLAST were limited and were unlikely to be representative of health utility during the entire post-relapse period. Instead, post-relapse utility estimates were based on an ATT matching analysis of the 63 BLAST PAS patients and patients recruited into the TOWER trial of blinatumomab in Ph- R/R BCP-ALL. The CS states that relapsed patients in the CR1 population of BLAST can be considered to be similar to patients in the TOWER trial who did not receive prior salvage therapy and who were not refractory at baseline. A utility value of 0.692 was estimated using this approach. The ERG notes that the precise methods used to generate this value are unclear due to the limited reporting in the CS and the redaction of utility estimates from the Appraisal Committee papers for TA450.[53]

HRQoL decrements associated with HSCT were based on a cross-sectional survey of 524 patients with acute leukaemia (75% acute myeloid leukaemia [AML], 25% ALL) in Japan (Kurosawa *et al*[51]). All patients undergoing HSCT are assumed to experience utility decrements of 0.17, 0.01, and 0.02 during years 1, 2, and 3-5 after HSCT, respectively, based on the differences in the mean utility value at these time points versus >5 years post-HSCT reported by Kurosawa *et al*. The company's model assumes that no further transplant-specific HRQoL decrement is applied 5-years post-HSCT.

A further HRQoL decrement of 0.02 is applied to the general population health utility values to reflect long-term effects of exposure to radiotherapy, chemotherapy, and HSCT. The CS states that this decrement was based on half the difference between the average utility value for blinatumomab patients in the RFS state, off therapy, and with MRD response (0.842) and the age- and sex-weighted mean population norms for patients between the ages of 35 and 55 (0.877).

Table 30 summarises the health utility values included in the company's model.

Table 30: Health state utilities applied in company's model

Health state	Utility	95% CI	Derivation
Blinatumomab, on-treatment, relapse-free, >6 months prior to death, cycle 1†	0.792	(0.699, 0.886)	Sampling of utility coefficients from the GLM/GEE model, the MRD response rate and baseline utility from the 63 BLAST PAS patients with data
Blinatumomab, on-treatment, relapse-free, >6 months prior to death, cycle 2+†	0.832	(0.789, 0.872)	
Blinatumomab, off-treatment, relapse-free, >6 months prior to death, cycle 1†	0.802	(0.708, 0898)	
Blinatumomab, off-treatment, relapse-free, >6 months prior to death, cycle 2+†	0.842	(0.798, 0.883)	
Standard care, relapse-free, >6 months prior to death	0.806	(0.718, 0.895)	
Blinatumomab and standard care, post-relapse >6 months prior to death	0.692	(0.688, 0.695)	Estimated from logistic regression of TOWER and the 63 BLAST PAS patients with data
General population utility decrement*	-0.02	N/a (constant)	Based on mid-point between utility from BLAST for RFS off-treatment, with MRD response and age- and sex-weighted general population norms[50]
HSCT utility decrement 1-12 months	-0.170	(-0.366, 0.026)	Estimated based on difference in utility from >5 years post-transplant and prior timepoints[51]
HSCT utility decrement 13-24 months	-0.010	(-0.096, 0.076)	
HSCT utility decrement 25-60 months	-0.020	(-0.085, 0.045)	
HSCT utility decrement 61 months+	0.000	N/a - constant	Assumption

** Decrement applied to all age-adjusted utility values*
† CrI generated by the ERG using the company's model
GLM/GEE – generalised linear model/generalised estimating equation; CI – confidence interval

Mean blinatumomab acquisition

Drug acquisition costs for blinatumomab were provided by the company. The company has a Patient Access Scheme in place for blinatumomab resulting in a price of [redacted] for one 38.5µg vial. The model assumes that one vial includes a single dose of useable medication (28µg blinatumomab). The model assumes that patients receive up to four cycles of blinatumomab at a mean dose of 28µg per day for 28 days, followed by 14 days off treatment. This dosing schedule is based on the anticipated marketing authorisation for blinatumomab,[24] rather than the dose used in BLAST (15µg/m^2).[1] Within the BLAST PAS, the mean body surface area (BSA) was 1.89m^2 which leads to a mean dose of 28.4µg, hence there is no difference in cost between the regimen used in BLAST and the regimen indicated by the marketing authorisation. The model estimates the costs of blinatumomab during each cycle using data on the average of the proportion of patients starting and completing each treatment cycle (see Table 31).

Table 31: Estimated percentage of patients starting and completing each cycle of blinatumomab

Blinatumomab treatment cycle	Patients starting cycle	Patients completing cycle	Assumed treatment proportion in each cycle
1	[redacted]	[redacted]	[redacted]
2	[redacted]	[redacted]	[redacted]
3	[redacted]	[redacted]	[redacted]
4	[redacted]	[redacted]	[redacted]

Blinatumomab administration and associated costs

The model assumes that the administration of blinatumomab is associated with costs relating to inpatient infusions, the pump used to deliver blinatumomab, and outpatient appointments to change the pump bag when treatment is delivered in a home setting.

During the first and second cycles of blinatumomab treatment, patients are assumed to receive 4 days and 2 days of inpatient treatment, respectively; no inpatient days are assumed to be spent delivering blinatumomab during cycles 3 or 4. The cost per day of administering blinatumomab in an inpatient setting was estimated to be £685.56. This value was based on NHS Reference Costs 2015/16[56] and was calculated as the finished consultant episodes (FCE) weighted average of unit costs divided by mean inpatient days for currency codes SA24G-J.

The pump used to deliver blinatumomab was estimated to cost £1,795 and was assumed to have a lifespan of 5 years. The daily cost of the pump was calculated assuming that the pump was used every day during its lifespan. An additional annual maintenance cost of £90 was assumed.

It was assumed that patients require an outpatient visit to change the bag in the pump every 4 days spent receiving blinatumomab in the outpatient setting. These visits were assumed to cost £211.99 per visit, based on NHS Reference Costs 2015/16[56] (outpatient, currency code SB15Z).

Costs associated with prophylaxis against CNS relapse given alongside blinatumomab

In the blinatumomab group, the model assumes that patients receive one outpatient visit per cycle to deliver prophylaxis against CNS relapse (methotrexate, cytarabine and dexamethasone) at a cost of £265.02 (derived from NHS Reference Costs 2015/16,[56] outpatient visit, code SB13Z - Deliver more

Complex Parenteral Chemotherapy at First Attendance). Table 32 summarises the prophylaxis acquisition costs applied in the company's model. For every regimen component, the dose was calculated based on the protocol and the mean BSA in the BLAST PAS. Costs were then calculated assuming that vials and tablets would be perfectly split.

Table 32: Costs associated with prophylaxis against CNS relapse included in the company's model

Treatment	Administration method	Unit size	Tablet/ vial size	Unit cost	Source
Methotrexate	Intrathecal	1000mg	1	£6.63	CMU[52]
Cytarabine	Intrathecal	2000mg	1	£6.60	CMU[52]
Dexamethasone	Intrathecal	3.3mg	10	£2.42	CMU[52]

CMU – Commercial Medicines Unit; mg - milligram

Standard care chemotherapy acquisition

Drug acquisition costs for the standard care group are summarised in Table 33. Standard care chemotherapy was assumed to follow the maintenance regimen for the non-transplanted population of the UKALL14 trial.[19] This regimen is assumed to be discontinued upon receipt of HSCT. Unit costs for all therapies were taken from the Commercial Medicines Unit (CMU) Electronic Marketing Information Tool (eMIT).[52] The model assumes vial sharing with no wastage of pills for oral treatments. For each regimen component, the dose was calculated based on the UKALL14 protocol and the mean BSA in the BLAST PAS. Costs were then calculated assuming that vials and tablets would be perfectly split.

Table 33: Drug acquisition costs applied in the standard care group

Treatment	Administration method	Unit size	Tablet/ vial size	Unit cost	Source
Vincristine	IV	2.0mg	5	£29.26	CMU[52]
Prednisolone	Oral	5mg	28	£0.41	CMU[52]
Mercaptopurine	Oral	50mg	25	£49.15	BNF[25]
Methotrexate	Oral	2.5mg	100	£4.39	CMU[52]
Methotrexate	Intrathecal	1000mg	1	£6.63	CMU[52]

CMU - Commercial Medicines Unit; BNF – British National Formulary; mg - milligram

Standard care chemotherapy administration

In the standard care group, the model assumes that patients receive two outpatient visits per cycle for IV administration of vincristine and intrathecal administration of methotrexate. For intrathecal methotrexate, the cost of administration was assumed to be £265.02, based on NHS Reference Costs 2015/16[56] (outpatient visit, code SB13Z - Deliver more Complex Parenteral Chemotherapy at First Attendance). For vincristine, the cost of administration was assumed to be £304.30, again based on NHS Reference Costs 2015/16[56] (outpatient visit, code SB14Z - Deliver Complex Chemotherapy, including Prolonged Infusional Treatment, at First Attendance). Patients are assumed to self-administer the oral components of the regimen.

Resource use associated with standard care and discontinued blinatumomab

The mean number of additional inpatient and outpatient visits (over and above drug administration visits) for patients who receive standard care and for those who discontinue blinatumomab (according to MRD response status) are based on estimates from face-to-face interviews with two UK experts.[1] These resource use estimates were combined with arm-specific MRD response rates. The MRD response rate in the blinatumomab arm (83.6%) was taken from the subgroup of the BLAST PAS who had an MRD response within the first two cycles of blinatumomab treatment. No data were available on delayed MRD response in the standard care group; the company's model assumes that MRD response for patients receiving standard care is 8.0% based on expert advice that *"this proportion is no greater than 10%"* (CS,[1] page 95). The resource use estimates applied in each treatment group are summarised in Table 34. The costs presented in this table are not applied to patients who are still receiving blinatumomab.

Table 34: Inpatient and outpatient resource use per month by MRD status and associated monthly resource use

Services	Face-to-face interview (n=2)		Resource use applied in each treatment group		Unit cost
	MRD +	MRD-	Discontinued blinatumomab	Standard care	
Inpatient days	1.75	0.06	0.337	1.615	£685.56
Haematologist - outpatient	2.000	1.500	1.918	1.960	£166.03[a]
Radiologist – outpatient	0.417	0.250	0.390	0.404	£51.35[b]
Other specialist – outpatient	0.500	0.250	0.459	0.480	£162.84[c]
General physician - outpatient	0.750	0.417	0.695	0.723	£36[d]

a NHS Reference Costs 2015/16,[56] consultant led face-to-face follow up. Currency code WF01A. Service code 303
b NHS Reference Costs 2015/16,[56] consultant led face to face follow up. Currency code WF01A. Service code 812
c NHS Reference Costs 2015/16,[56] consultant led face to face follow up. Currency code WF01A. Service code 370
d Curtis and Burns 2016
MRD+ - molecular evidence of blasts in the bone above 1 in 10,000, MRD- molecular evidence of blasts in the bone below 1 in 10,000,

Costs associated with HSCT

The model assumes that patients may receive HSCT prior to relapse and/or following relapse. The company's model assumes that patients who are relapse-free may undergo HSCT for up to four years after initiation of treatment with blinatumomab or standard care chemotherapy. The model uses data on the cumulative 4-year probability of having undergone pre-relapse HSCT from the BLAST PAS ([redacted]) and the ATT-weighted historical control DCAS ([redacted]) to inform the blinatumomab and standard care groups, respectively. The modelled 6-monthly probability of receiving HSCT was calibrated such that the predicted cumulative probability of having undergone pre-relapse HSCT at 48 months matches the observed cumulative probabilities.

In the post-relapse population, the model uses four probability inputs as well as the treatment-specific RFS curve to determine the per-cycle probability of receiving post-relapse HSCT. Two probabilities are used to estimate the probability of receiving a post-relapse HSCT: (i) the probability of having a post-relapse HSCT conditional on the patient not having had a pre-relapse HSCT (probability=0.20); (ii) the probability of having a post-relapse HSCT conditional on the patient having previously had a pre-relapse HSCT (probability=0.16). The exact methods and evidence used to estimate these parameters are not clear from the CS.[1] The remaining two probabilities relate to the probability that an RFS event is death in each treatment group (as described in Table 29). These probabilities were estimated from the BLAST PAS for the blinatumomab arm and the historical control DCAS for the standard care arm.

The model predictions for the mean number of HSCTs per patient are summarised in Table 35. As shown in the table, the company's model suggests that the mean number of HSCTs is higher in the blinatumomab group than the standard care group (mean HSCTs blinatumomab versus standard care - 0.79 versus [redacted]).

Table 35: Mean number of HSCTs per patient predicted by the company's model

Treatment group	Mean number of HSCTs per patient		
	Pre-relapse	**Post-relapse**	**Total**
Blinatumomab	[redacted]	[redacted]	[redacted]
Standard care	[redacted]	[redacted]	[redacted]

Salvage chemotherapy costs

The salvage chemotherapy regimen is assumed to be FLAG-IDA. The cost of this regimen was estimated to be £16,175 (uplifted to 2015/16 prices), based on the cost estimates reported in NICE TA450.[53] The model assumes that 37% of patients who receive one line of salvage therapy also receive a second line of salvage therapy; this results in a total cost of £21,905 per patient receiving salvage therapy.

Terminal care costs

The model assumes that at the end of life, patients spend 8 weeks (56 days) receiving hospital care. The cost of care (uplifted to 2015/16 prices) was estimated to be £157.74 per day.[55] The mean cost of terminal care was estimated to be £8,834 per patient.

Model evaluation methods

The CS presents the results of the economic analysis in terms of the incremental cost per QALY gained for blinatumomab versus standard care. Results are presented for both the deterministic and probabilistic versions of the model. The CS also includes the results of probabilistic sensitivity analysis (PSA), deterministic sensitivity analyses (DSAs) and scenario analyses. The results of the PSA are presented in the form of a cost-effectiveness plane and cost-effectiveness acceptability curves (CEACs), based on 10,000 Monte Carlo simulations. The results of the DSAs are presented in the form of a tornado diagram for specified model parameters (based on their 95% confidence limits). Alternative scenario analyses are also reported to explore the use of ATE weights, alternative choices of RFS and OS curves, alternative assumptions regarding long-term excess mortality, duration of blinatumomab benefit, and alternative assumptions regarding the probability that an RFS event is death, HSCT use, probability of cure, HRQoL, costs, discount rates and the model time horizon. The distributions applied in the company's PSA are summarised in Table 36. The ERG notes that several uncertain parameters are held fixed at their mean values and some of the choices of distribution and derived standard errors are not appropriate.

Table 36: Distributions applied in company's probabilistic sensitivity analyses

Parameter type	Parameter	Distribution	ERG comment
Patient characteristics	Age, sex, BSA and weight	Fixed	-
Time-to-event parameters	RFS – blinatumomab	Bootstrap	No details provided regarding how the bootstrap procedure was undertaken. It is unclear whether uncertainty in the IPTW weights was included
	RFS – standard of care	Bootstrap	
	OS – blinatumomab	Bootstrap	
	OS – standard of care	Bootstrap	
Probability RFS event is death	RFS death probability - blinatumomab	Beta	-
	RFS death probability – standard care	Beta	-
HRQoL	RFS health utility model baseline	Log normal	Distribution is not bounded by zero and 1.0
	RFS health utility GLM/GEE model parameters (intercept, baseline, off-treatment relapse-free, MRD response and terminal decrement)	Multivariate normal	-
	Health utility - relapsed	Log normal	Distribution is not bounded by zero and 1.0
	QALY loss (time-dependent) – HSCT	Normal	Distribution is not bounded by zero. The HRQoL decrements for HSCT includes positive values in the PSA; this is illogical.
	General population utilities	Fixed	These values are subject to uncertainty
	Utility decrement – exposure to radiotherapy, chemotherapy, and HSCT	Fixed	These values are subject to uncertainty
Mean dosing	Proportion of full blinatumomab dose received (up to 4 doses)	Beta	-
Probability of receiving HSCT	Probability of receiving HSCT pre-relapse	Beta	The per-cycle probability, rather than the 4-yearly probability has been included in the PSA
	Probabilities of receiving HSCT post-relapse	Beta	-
Resource use and costs	Inpatient and outpatient resource use for standard of care and discontinued blinatumomab	Fixed	These values are subject to uncertainty
	HSCT procedure and subsequent follow-up (0-24 months)	Log normal	SE arbitrarily assumed to be 25% of mean
	Maintenance chemotherapy (standard care group)	Fixed	SE arbitrarily assumed to be 25% of mean
	Salvage chemotherapy	Log normal	SE arbitrarily assumed to be 25% of mean
	Terminal care costs	Log normal	SE arbitrarily assumed to be 25% of mean

Parameter type	Parameter	Distribution	ERG comment
	Prophylaxis against CNS relapse for patients receiving blinatumomab	Fixed	-
	Blinatumomab acquisition cost (including PAS)	Fixed	-
	Cost of pump use to deliver blinatumomab	Log normal	SE arbitrarily assumed to be 25% of mean
	Cost of inpatient visits for blinatumomab	Log normal	SE arbitrarily assumed to be 25% of mean
	Unit costs for visits, appointments, hospitalisations, laboratory tests, radiological tests and AEs	Log normal	SE arbitrarily assumed to be 25% of mean. Given that these are based on NHS Reference Costs, SEs could have been calculated using reported interquartile ranges.

SE – standard error

Company's model verification and validation methods

The CS[1] details extensive efforts taken to verify the correct implementation of the model and to ensure the accuracy of the model inputs against the source material from which these were derived. The CS[1] and the clarification response[17] also mention the use of clinical experts to inform certain assumptions within the model (e.g. around the plausibility of cure).

Company's model results

Table 37 presents the central estimates of cost-effectiveness derived from the company's model. Based on the probabilistic version of the model (assuming the unrestricted Gompertz function for RFS and the log normal mixture cure model for OS), blinatumomab is expected to generate an additional 2.85 QALYs at an additional cost of £84,456 compared with standard care: the corresponding incremental cost-effectiveness ratio (ICER) for blinatumomab versus standard care is £29,673 per QALY gained. The deterministic version of the company's model produces a similar ICER of £28,524 per QALY gained for blinatumomab versus standard care.

Table 37: Company's base case cost-effectiveness results – blinatumomab versus standard care (original submitted model)

Probabilistic model					
Option	**QALYs**	**Costs**	**Inc. QALYs**	**Inc. costs**	**Incremental cost per QALY gained**
Blinatumomab	6.96	[redacted]	2.85	£84,456	**£29,673**
Standard care	4.11	[redacted]	-	-	-
Deterministic model					
Option	**QALYs**	**Costs**	**Inc. QALYs**	**Inc. costs**	**Incremental cost per QALY gained**
Blinatumomab	7.10	[redacted]	2.95	£84,259	**£28,524**
Standard care	4.14	[redacted]	-	-	-

Figure 9 and Figure 10 present the results of the company's PSA in the form of a cost-effectiveness plane and CEACs, based on a re-run of the company's original submitted model. Assuming a willingness-to-pay (WTP) threshold (λ) of £20,000 per QALY gained, the company's model suggests that the probability that blinatumomab produces more net benefit than standard care is 0.10. Assuming a WTP threshold of £30,000 per QALY gained, the probability that blinatumomab produces more net benefit than standard care is estimated to be 0.53.

Figure 9: Cost-effectiveness plane – blinatumomab versus standard care (adapted from company's model)

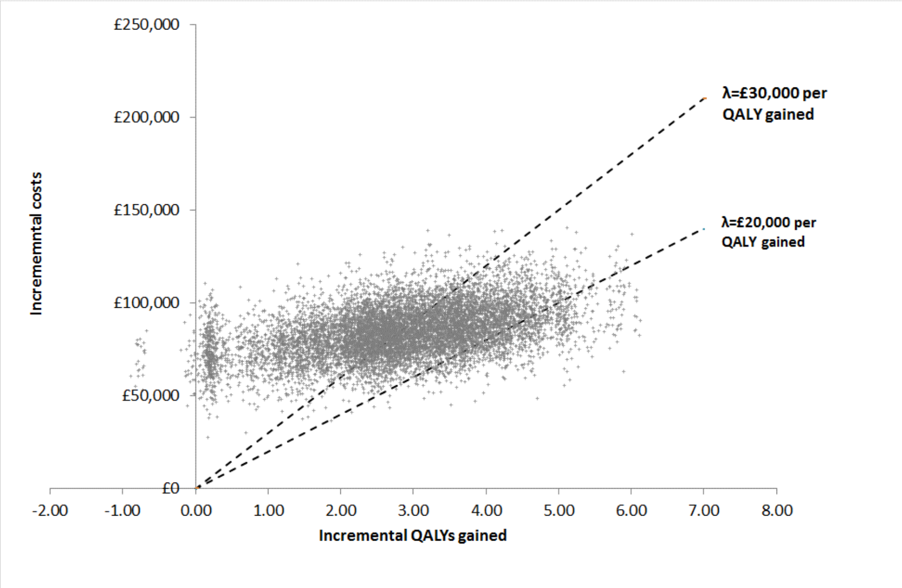


Figure 10: Cost-effectiveness acceptability curves – blinatumomab versus standard care (adapted from company's model)

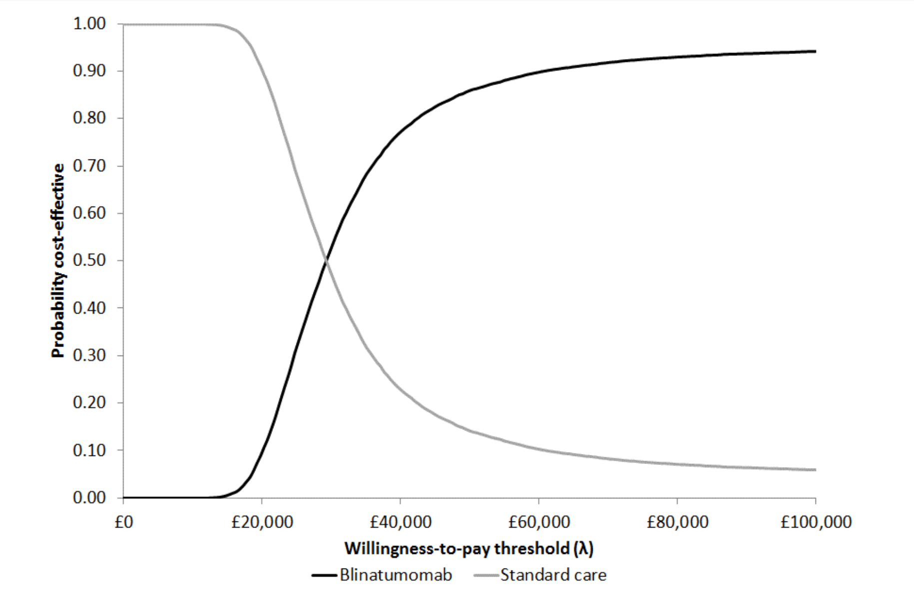


Figure 11 presents the results of the company's DSAs. The DSAs indicate that the five most influential model parameters relate to: (i) the blinatumomab OS cure fraction; (ii) the standard care OS cure fraction; (iii) the proportion of patients in the blinatumomab group who receive HSCT; (iv) the duration of blinatumomab therapy, and (v) the cost of HSCT.

Figure 11: Deterministic sensitivity analysis results – blinatumomab versus standard care (reproduced from company's model)

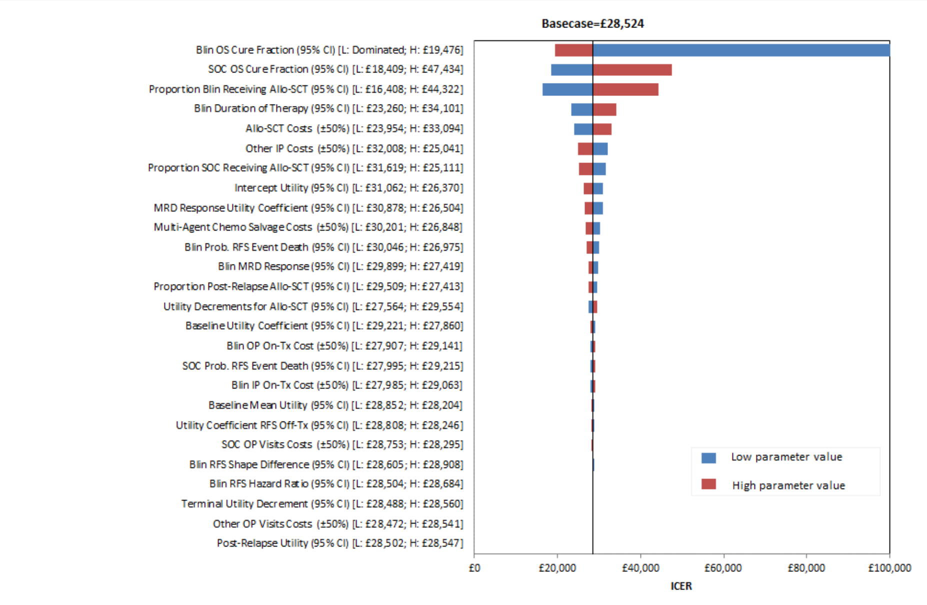


Table 38 presents the results of the company's scenario analyses. The ICER for blinatumomab versus standard care appears to be generally robust to most of the scenarios tested, although the ICER is greater than £30,000 per QALY gained for many scenarios tested. The ICER appears to be particularly influenced by the use of ATE weights, the duration of therapy, the use of the health care resource use survey, inflating OS and RFS outcomes in both groups and the cure fraction.

Table 38: Scenario analysis results – blinatumomab versus standard care (adapted from CS Table 62)

No.	Scenario	Inc. QALYs	Inc. costs	ICER
1	Base case	2.95	£84,259	**£28,524**
1	ATE weights	2.39	£81,370	**£33,999**
2	Alternative extrapolation methods	3.31	£83,064	**£25,081**
3	Unfavourable - RFS RCS log logistic (R), OS RCS Weibull (R)	2.74	£83,874	**£30,647**
4	2-fold increase long-term excess mortality	3.35	£84,300	**£25,199**
5	6-fold increase long-term excess mortality	2.69	£84,234	**£31,274**
6	Duration of benefits = 60 months	2.44	£84,263	**£34,559**
7	Inpatient costs with on-treatment inpatient days from BLAST	2.95	£89,235	**£30,209**
8	Inpatient costs with on-treatment inpatient days from blinatumomab label	2.95	£84,405	**£28,574**
9	23.55% of blinatumomab RFS events are deaths	2.95	£90,548	**£30,698**
10	HRU data from online survey	2.95	£105,376	**£35,673**
11	Cumulative probability of pre-relapse HSCT same for blinatumomab as for standard care	3.00	£49,403	**£16,479**
12	ALL-related costs applied to end of model time horizon	2.95	£80,302	**£27,185**
13	0% MRD response rate for standard care	2.96	£82,537	**£27,892**
14	15% MRD response rate for standard care	2.95	£85,766	**£29,080**
15	No disutility for long-term survivors	3.01	£84,259	**£27,979**
16	0.04 disutility for long-term survivors	2.90	£84,259	**£29,091**
17	Standard care RFS utility = blinatumomab off-treatment RFS utility	2.93	£84,259	**£28,722**
18	ALL-related utilities and costs only to 36 months	2.92	£87,100	**£29,866**
19	ALL-related utilities and costs only to 48 months	2.94	£85,364	**£29,056**
20	Model timeframe = 30 y	2.85	£84,126	**£29,552**
21	Model timeframe = 60 y	2.95	£84,259	**£28,524**
22	Annual discount rate for costs and QALYs=1.5%	3.76	£85,119	**£22,639**
23	Limitations relating to generalisability of standard care arm to current practice (RFS and OS survival distribution based on the ATT analysis of the historical cohort study is adjusted upwards by a factor of 15%)	2.22	£80,202	**£36,163**
24	Blinatumomab OS cure fraction = midpoint OS cure fractions (incremental cure fraction halved)	1.61	£78,918	**£49,101**

ATE – average treatment effect; ATT – average treatment effect on the treated; R – restricted; RCS – restricted cubic spline; HRU – health care resource use

Updated model results

In response to minor issues raised by the ERG during the clarification process, the company provided an updated model which included the following amendments: (i) maximum annual mortality risk capped at 100%; (ii) pump costs included for all days after the first inpatient stay; (iii) general population utilities based on Ara and Brazier,[58] and (iv) post-relapse allogeneic HSCT not initiated after 5 years. The updated model results are similar to the company's original base case (see Table 39); the probabilistic ICER for blinatumomab versus standard care is estimated to be £28,655 per QALY gained.

Table 39: Company's base case cost-effectiveness results – blinatumomab versus standard care (updated model submitted following clarification)

Probabilistic model					
Option	**QALYs**	**Costs**	**Inc. QALYs**	**Inc. Costs**	**Incremental cost per QALY gained**
Blinatumomab	7.11	[redacted]	2.92	£83,634	**£28,655**
Standard care	4.19	[redacted]	-	-	-
Deterministic model					
Option	**QALYs**	**Costs**	**Inc. QALYs**	**Inc. Costs**	**Incremental cost per QALY gained**
Blinatumomab	7.23	[redacted]	3.02	£83,800	**£27,779**
Standard care	4.21	[redacted]	-	-	-

5.3 Critical appraisal of the company's health economic analysis

This section presents a critical appraisal of the health economic analysis presented within the CS.[1] Section 5.3.1 details the methods used by the ERG to interrogate and critically appraise the company's submitted health economic analysis. Section 5.3.2 discusses the extent to which the company's analysis adheres to the NICE Reference Case.[26] Section 5.3.3 summarises the ERG's verification of the company's implemented model and highlights inconsistencies between the model, the CS, and the sources used to inform the model parameter values. Section 5.3.4 presents a detailed critique of the main issues and concerns underlying the company's analysis.

5.3.1 Methods for reviewing the company's economic evaluation and health economic model

The ERG adopted a number of approaches to explore, interrogate and critically appraise the company's submitted economic evaluation and the underlying health economic model upon which this was based. These included:

- Consideration of key items contained within published economic evaluation and health economic modelling checklists[59, 60] to critically appraise the company's model and analysis.
- Scrutiny of the company's model by health economic modellers and discussion of issues identified amongst the members of the ERG.
- Double-programming of the deterministic version of the company's model to fully assess the logic of the model structure, to draw out any unwritten assumptions and to identify any apparent errors in the implementation of the model.
- Examination of the correspondence between the description of the model reported within the CS[1] and the company's executable model.
- Replication of the base case results and PSA presented within the CS.[1]
- Where possible, checking of parameter values used in the company's model against their original data sources.

- The use of expert clinical input to judge the credibility of the company's economic evaluation and the assumptions underpinning the model.

5.3.2 Adherence of the company's model to the NICE Reference Case

The company's economic evaluation is generally in line with the NICE Reference Case[26] (see Table 40). The ERG notes that the model excludes relevant patient subgroups which are included in the proposed marketing authorisation and that inevitably there is considerable uncertainty surrounding the results of the analysis due to the observational nature of the data. These issues are discussed in further detail in Section 5.3.4.

Table 40: Adherence of the company's model to the NICE Reference Case

Element	Reference case	ERG comments
Defining the decision problem	The scope developed by NICE	The ERG notes that the model reflects a population of patients who are able to receive chemotherapy; however, blinatumomab represents a potential treatment option for patients who are unable to undergo HSCT or to tolerate chemotherapy. In addition, two further potentially overlapping subgroups of the BLAST study were excluded from the indirect comparison and health economic model: (i) patients with Ph- MRD+ BCP-ALL in CR2+; (ii) patients with Ph+ MRD+ BCP ALL.
Comparator(s)	As listed in the scope developed by NICE	The company's model compares blinatumomab against standard care chemotherapy. The final NICE scope[22] included a second comparator which was defined as "monitor for relapse." This has not been included as an option in the company's model; the ERG notes that this comparator would be relevant to patients who are unable to receive HSCT or to tolerate chemotherapy.
Perspective on outcomes	All direct health effects, whether for patients or, when relevant, carers	Health gains accrued by patients are modelled in terms of QALYs gained.
Perspective on costs	NHS and PSS	Whilst not explicitly stated in the CS,[1] the company's economic analysis adopts an NHS and PSS perspective.
Type of economic evaluation	Cost-utility analysis with fully incremental analysis	The company's economic evaluation takes the form of a cost-utility analysis. The results of the analysis are presented in terms of the incremental cost per QALY gained for blinatumomab versus standard care.
Time horizon	Long enough to reflect all important differences in costs or outcomes between the technologies being compared	The company's model adopts a 50-year time horizon. By this timepoint, more than 99.9% of the modelled population have died.
Synthesis of evidence on health effects	Based on systematic review	Health outcomes are modelled using IPTW weighted data from the BLAST PAS and the historical control DCAS (both studies are currently unpublished).

Element	Reference case	ERG comments
Measuring and valuing health effects	Health effects should be expressed in QALYs. The EQ-5D is the preferred measure of HRQoL in adults.	HRQoL estimates for the relapse-free state were derived from GLM/GEE regression analyses of patient-reported EQ-5D data collected in the BLAST study.[1] The HRQoL estimate for the post-relapse state was derived from a logistic regression analysis using the TOWER trial and the 63 patients in the BLAST study with HRQoL data.[1] Additional HRQoL estimates are based on the literature[50, 51] and assumptions.
Source of data for measurement of health-related quality of life	Reported directly by patients and/or carers	
Source of preference data for valuation of changes in HRQoL	Representative sample of the UK population	
Equity considerations	An additional QALY has the same weight regardless of the other characteristics of the individuals receiving the health benefit	No additional equity weighting is applied to estimated QALY gains. The CS argues that blinatumomab meets NICE's criteria for a life-extending end of life treatment. The CS also argues that blinatumomab meets many of the criteria for appraisal under the NICE HST framework and should be evaluated taking into account a wider range of criteria about the benefits and costs.
Evidence on resource use and costs	Costs should relate to NHS and PSS resources and should be valued using the prices relevant to the NHS and PSS	Resource components included in the company's model reflect those relevant to the NHS and PSS. Unit costs were valued at 2015/16 prices.
Discount rate	The same annual rate for both costs and health effects (currently 3.5%)	Costs and health effects are discounted at a rate of 3.5% per annum.

PSS – Personal Social Services; HRQoL – health-related quality of life

5.3.3 Model verification and correspondence between the model, the CS and parameter sources

Model verification

The ERG rebuilt the deterministic version of the company's base case model in order to verify its implementation; the results of the model rebuild are shown in Table 41. As shown in the table, the ERG's rebuilt model produces very similar estimates of health gains, costs and cost-effectiveness compared with the company's model. During the process of rebuilding the company's base case economic model, seven minor implementation/programming errors were identified:

(i) The annual general population mortality rate is applied for 1-year intervals defined according to time since model entry, rather than according to patient age. However, the initial patient age is not an integer (initial age = 45.38 years), hence applying the modelled mortality =LOOKUP() function for a full year is incorrect.

(ii) The risk of all-cause death exceeds 1.0 for males patients aged 95 years and older and female patients aged 97 years and older. This error was rectified within the company's updated model provided as part of the company's clarification response[17] (question B42).

(iii) The formula used to calculate the receipt of HSCT at 2 years is subject to minor programming errors.

(iv) The formula used to apply discounting to the cost of other inpatient visits post-relapse in the blinatumomab arm is subject to a programming error caused by the formula being incorrectly offset.

(v) Post-relapse HSCTs were assumed to occur after the 5-year time point; however, elsewhere in the model, ALL-related costs were not applied after 5 years as it was assumed that patients would not relapse beyond this timepoint. This error was rectified within the company's updated model provided as part of the company's clarification response[17] (question B33).

(vi) The application of the utility decrement due to death for each model cycle was calculated based on the number of deaths occurring 6 months into the future. Within the first 5 years of the time horizon, the model assumes that all deaths are ALL-related and should therefore be subject to the utility decrement (based on the GLM/GEE model). The model multiplies the utility decrement calculated from the GLM/GEE by the number of people who were expected to die either within either: (i) the next 27 model cycles, or (ii) before the model time horizon reaches 5 years. This approach is inappropriate, as the utility decrement for a patient who dies within a model time cycle should depend on the patient's current survival probability and their history, rather than events occurring in the future.

(vii) Discounting is incorrectly applied to the HSCT costs due to the use of approximate =LOOKUP() functions used to calculate the discount rate for receipt of HSCT.

The ERG notes that these errors have only a minor impact on the ICER for blinatumomab versus standard care.

Table 41: Comparison of company's original submitted base case model and ERG's rebuilt model including PAS

Option	Company's model			ERG's rebuilt model		
	QALYs	Costs	ICER	QALYs	Costs	ICER
Blinatumomab	7.10	[redacted]	**£28,524**	7.10	[redacted]	**£28,529**
Standard care	4.14	[redacted]	-	4.14	[redacted]	-

QALY – quality-adjusted life year; ICER - incremental cost-effectiveness ratio

Correspondence between the written submission and the model

The implemented model appears to be generally in line with its description within the CS.[1] However, the ERG considers that the logic and implementation of the HSCT sub-models are not well described in the CS. In addition, limited detail is provided regarding the logistic regression of the TOWER and BLAST data used to generate the post-relapse utility value. As individual patient-level data (IPD) were not provided by the company, it was not possible for the ERG to fully verify the implementation of the survival models described in the CS.

Correspondence of the model inputs and the original sources of parameter values

The ERG was unable to locate the company's estimated cost of death within the King's Fund and Marie Curie reports;[54, 55] however, the value used within the model should not have a material impact on the model results. In addition, the ERG could not identify the cost of salvage therapy (£16,175) or the source of the assumption that 37% of patients receiving salvage therapy would receive a subsequent further line of salvage therapy within the Appraisal Committee papers from TA450.[53] As the company produced these analyses for an earlier appraisal, this lack of correspondence is unlikely to be an important issue. Further, the ERG was unable to source the parameter value relating to the proportion of patients who survive 24 months after receiving HSCT (20%). All other parameter values correspond with their original sources.

5.3.4 Main issues identified within the critical appraisal

Box 1 summarises the main issues identified within the ERG's critical appraisal of the company's economic analysis. These issues are discussed in further detail in the subsequent sections.

Box 1: Summary of main issues identified within the company's health economic model

(1) Exclusion of relevant patient groups from the economic analysis
(2) "Monitor for relapse" comparator not included in the model
(3) Use of a model structure which is inappropriate for tracking HSCT
(4) Absence of RCT evidence for blinatumomab versus standard care
(5) Concerns regarding company's approach to RFS/OS model selection
(6) Concerns regarding the robustness of the company's alternative base case (blinatumomab used on relapse for the standard care group)
(7) Questionable reliability of the company's HRQoL estimates
(8) Uncertainty surrounding the proportion of RFS events that are deaths
(9) Unrealistic treatment pathway
(10) Limited sensitivity analysis around alternative parametric functions

(1) Exclusion of relevant patient groups from the economic analysis

The population considered within the company's economic analysis relates to patients with Ph- MRD+ BCP-ALL with first complete haematological remission (CR1). This modelled population is narrower than the anticipated marketing authorisation for blinatumomab,[24] as it excludes three relevant subgroups of patients: (i) patients who are unable to receive HSCT or to tolerate chemotherapy; (ii) patients with Ph- MRD+ BCP- ALL with CR2+, and (iii) patients with Ph+ MRD+ BCP-ALL. The CS[1] argues that blinatumomab should be considered for use in alignment with its full anticipated marketing authorisation (for the treatment of adults with MRD+ BCP-ALL).

In response to a request for clarification (see clarification response,[17] question A2), the company noted that there is limited evidence relating to patients with CR2+. Based on the results of the BLAST study, patients with CR1 and MRD response had better outcomes than patients with CR2 and MRD response, however, those in CR2 and MRD response still gained benefit from blinatumomab (see Table 42). However, the historical control study included only patients with CR1, hence there are no data available for comparison. The CS[1] and the company's clarification response[17] also noted that clinical advice received by the company suggested that blinatumomab would be used as early in the pathway as possible and that *"subsequent use of blinatumomab to treat MRD positivity in later remission states or as a salvage therapy is not anticipated if blinatumomab is used in the aforementioned [first-line] setting."* On this basis, the company argues that the CR1 population is the most appropriate ICER for decision-making. Clinical advisors to the ERG agreed that blinatumomab would be used as early as possible in the treatment pathway. However, the ERG notes that the exclusion of patients with CR2+ reduces the available sample size from the BLAST study (41 of 116 [35.3%] patients had second or third CR).

Table 42: Summary of OS and RFS for blinatumomab-treated patients in CR2 in BLAST (adapted from clarification response question A2)

CR2 subpopulation, BLAST	MRD responders	MRD non-responders
RFS, median (months)	[redacted]	[redacted]
OS, median (months)	[redacted]	[redacted]

The company's clarification response[17] notes that the Ph+ population was not represented in the model as the number of Ph+ patients recruited into BLAST was very small (n=5), and the historical control study did not include these patients. Clinical advisors to the ERG stated that the treatment pathway for Ph+ ALL is markedly different from that for Ph- ALL, as several effective treatment options (specifically, TKIs) are available for these patients.

The CS makes the argument that MRD+ patients who have a high risk of relapse would not solely be monitored for relapse without any active treatment. However, clinical advisors to the ERG suggested that some patients may not be sufficiently fit to receive HSCT or chemotherapy, but may be able to tolerate blinatumomab. The company's model does not assess the cost-effectiveness of blinatumomab within this population.

The ERG considers that on the basis of the evidence submitted to NICE, it is not possible to make any reliable estimate of the cost-effectiveness of blinatumomab in these excluded population groups.

(2) "Monitor for relapse" comparator not included in the model

The company's model compares blinatumomab against standard care chemotherapy. The final NICE scope[22] listed an additional comparator which was defined as "monitor for relapse"; this option is not considered as a comparator in the company's model (see critical appraisal point 1). The ERG considers that the company's economic analysis should have explored an assessment of the cost-effectiveness of blinatumomab versus monitoring within the subgroup of patients who are unable to receive HSCT or retreatment with chemotherapy, but for whom blinatumomab is an option.

(3) Issues relating to the modelling of HSCT

The model attempts to incorporate the impact of HSCT through two mechanisms: (i) the principal benefits of HSCT in reducing and/or avoiding the risk of relapse and death are implicitly reflected in the RFS and OS time-to-event analyses, and (ii) the QALY losses and costs associated with the HSCT procedure and post-HSCT survival are reflected within two HSCT sub-models. The approach adopted by the company to capture the impact of HSCT is subject to several limitations: (a) the absence of a causal link between HSCT uptake and its impact on RFS and OS outcomes; (b) the model cannot estimate the probability that a patient receives HSCT; (c) the adoption of questionable assumptions regarding HSCT receipt, and (d) the likely underestimation of post-HSCT costs.

(i) Absence of a causal link between HSCT uptake and its impact on RFS and OS outcomes

The model does not include a causal link between the extent of HSCT use and the principal RFS/OS benefits resulting from the use of this intervention. For example, setting the 6-monthly probability of receiving HSCT to zero reduces the HSCT-related costs and QALY losses to zero, however, the RFS and OS outcomes remain unchanged. The absence of a direct structural link between the extent of HSCT use and the benefits and costs accrued as a consequence of HSCT makes it difficult to judge the reliability of this aspect of the model.

(ii) The model cannot estimate the probability that a patient receives HSCT

Given that HSCT is not explicitly incorporated into the company's model structure, it is not possible to track the proportion of patients who undergo HSCT post-relapse (including those patients who undergo more than one transplant). As such, it is not possible to calculate the proportion of people who would receive an HSCT (although it is possible to estimate the overall number of HSCTs received per patient). As a consequence, this aspect of the model is not transparent and it is difficult to determine whether the assumed use of HSCT is clinically plausible.

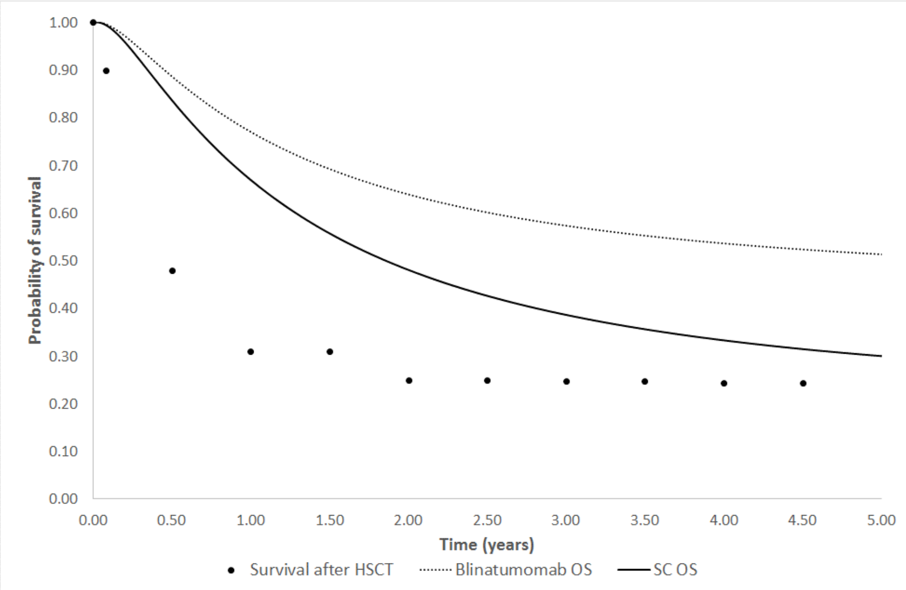
(iii) Adoption of questionable assumptions regarding HSCT receipt

The HSCT post-relapse sub-model makes the following assumptions: (i) HSCTs occur in 6-monthly batches (thereby affecting the cumulative proportion of patients still receiving chemotherapy at each timepoint), and (ii) patients with pre-relapse HSCT do not relapse until all patients without pre-relapse HSCT have relapsed. The first assumption could have been avoided by using the same cycle duration within the HSCT sub-models and the main partitioned survival model. The second assumption could have been avoided only through the use of a different overall model structure which would allow for the tracking of HSCT history across the patient cohorts. Whilst the data used to inform the frequency of post-relapse HSCT is weaker than that for pre-relapse HSCT, the overall frequency of post-relapse HSCT is low in both treatment groups, hence it does not have a substantial impact on the ICER.

(iv) Likely underestimation of post-HSCT costs

The probability of remaining alive post-HSCT over time is approximated using data on survival post-HSCT from NHS Blood and Transplant[48] (it is implicitly assumed that all surviving patients remain in post-HSCT follow-up) and using uplifted general population mortality estimates (from 2-years post-transplant onwards). These data are used to estimate the costs and health losses associated with HSCT and post-transplant care, but do not affect survival gains (see point (3i) above). Figure 12 presents a comparison of the parametric (log normal mixture cure) OS curves for each treatment group and the assumed survival post-HSCT applied in the HSCT sub-models. It should be noted that the modelled OS curves reflect what happens to all patients, including those who receive HSCT as well as those who do not, whilst the NHS Blood and Transplant data reflect survival in an exclusively transplanted cohort. Clinical advisors to the ERG noted that they would expect that, other things being equal, OS would be higher in transplanted patients compared to non-transplanted patients. However, as shown in Figure 12, OS in the transplanted cohort is markedly worse than that for both the blinatumomab and standard care groups. The consequence is that the model appears to include significant benefits in terms of OS due to cure following HSCT, but underestimates both the long-term costs and QALY losses associated with this treatment. Model testing undertaken by the ERG indicates that increasing the costs of post-HSCT follow-up and cyclosporine and increasing the HRQoL decrements associated with HSCT both lead to increases in the ICER for blinatumomab versus standard care.

Figure 12: Comparison of the assumed survival post-HSCT to the company's base case OS curves in both the blinatumomab and standard care arms for someone who received their HSCT within the first cycle of the HSCT sub-model



The ERG notes that in order to explicitly capture the extent of HSCT use pre- and post-relapse, and the costs and benefits accruing as a consequence of those procedures, a different model structure would be required (e.g. a semi-Markov model or a discrete event simulation [DES]). This would allow for tracking of patient histories, however, it would also require a re-analysis of the available time-to-event data to account for competing risks of relapse and death within transplanted and non-transplanted subgroups. The ERG believes that following such an approach would lead to two key benefits: (i) the incorporation of structural links between the use of HSCT and its associated costs and health impacts; (ii) the incorporation of more explicit assumptions regarding the benefits of HSCT (e.g. survival in transplanted and non-transplanted patients) which would improve model transparency and credibility. However, the ERG notes that the available data to populate specific transitions would be limited by very small sample sizes, may be subject to selection bias, and would be associated with considerable uncertainty.

(4) Absence of RCT evidence for blinatumomab versus standard care

As described in Section 4.4, propensity score methods based on IPTW were used to provide adjusted Kaplan-Meier survival curves for the standard care chemotherapy group. Although this is appropriate given the absence of RCT evidence, this introduces an important limitation for all subsequent analyses.

The propensity score weights (and hence the adjusted Kaplan-Meier survival curves) are estimates with associated uncertainty. It is unclear (although unlikely) that this uncertainty has been accounted for in the subsequent model fitting.

(5) Concerns regarding company's approach to RFS/OS model selection

The ERG has concerns regarding the company's approach to model selection. As detailed in Section 5.2, the company fitted a large number of parametric models to the available RFS and OS data. The company then selected the five best fitting RFS curves judged according to their BIC statistics and the five best fitting OS curves judged according to their BIC statistics and whether the given OS model was logically consistent with the final selected RFS function. Other aspects of model choice were considered only for the five best-fitting functions. The ERG notes that the CS does not provide any information regarding the use of clinical judgement to inform decisions regarding the plausibility of the selected RFS and OS functions, however, the company's clarification response[17] (question B7) states that UK clinicians were asked to comment on: (i) the expected survival of patients currently observed in clinical practice (at landmark timepoints); (ii) the appropriateness of assuming a cure at a specific timepoint; (iii) the proportion of patients that may realise a cure given current treatments, and (iv) the magnitude of benefit likely to be derived from obtaining an MRD-negative status. On the basis of the information provided in the CS, it does not appear that clinicians were asked to judge which specific parametric models appear most plausible.

The ERG considers that many of the curves fitted by the company are unnecessary and/or inappropriate. Clinical advice received by the company (see CS[1] page 120 and clarification response[17] question B7) and the ERG suggests that patients who have not relapsed within 5-years may generally be considered to be cured. Therefore, it seems reasonable to assume that a cure model is appropriate from the outset. However, the company's use of BIC to select out a subset of models for further consideration results in a situation whereby two of the "best" RFS models do not predict a cure fraction and are thus clearly inappropriate. It may be the case that other models which fit the data less well during the observed period may produce more plausible extrapolations, however these are excluded from further consideration due to company's application of an initial model selection criterion based on BIC.

The company fitted restricted, unrestricted, cure, cure (restricted) and cure (unrestricted) models to the available time-to-event data. The ERG considers that it would be appropriate to include only unrestricted models from the outset for two reasons. Firstly, whilst it is possible to explore the assumption of proportional hazards/constant accelerated failure over the observed period of the studies included in the analyses, this assumption may not hold within the extrapolated period, hence, the ERG would prefer to exclude models which apply such restrictive assumptions. Secondly, the data for the comparator are weighted but not directly observed and are subject to uncertainty. This uncertainty appears to have been ignored in the analysis presented by the company, with equal consideration given to the observed BLAST data and the weighted historical control data, with the latter having a larger sample size and so having a higher influence in the resulting model fit statistics. It would therefore be more appropriate to conduct the model fitting separately in both groups.

The ERG considers that these choices around the use of a cure model and the use of treatment effect covariates should have been made *a priori*. The ERG also notes that the company's model selection should have explored the clinical plausibility of the fitted models (based on full models which include the mortality hazard for cured patients).

In addition, the ERG notes that there appears to be some inconsistency between the clinical advice received by the company regarding the likelihood of achieving cure and the time at which the model predicts that such cure occurs. Figure 13 and Figure 14 present the modelled RFS and OS functions for the blinatumomab and standard care groups, respectively. The crosses marked on each RFS/OS curve show the "cure point" predicted by the model, that is, the timepoint at which the hazard for RFS or OS drops below the company's assumed hazard of other-cause mortality within this population. Beyond this timepoint, the model assumes that the only remaining event is other-cause death (uplifted from general population life tables). As shown in both Figure 13 and Figure 14, the timepoint at which the modelled RFS/OS event hazard reverts to that for the uplifted general population mortality differs between the RFS and OS endpoints within the same population. For the blinatumomab group, the model indicates a cure point at 7.28 years for RFS and 8.01 years for OS. Within the standard care group, the difference between the RFS and OS cure points is more pronounced, with cure being modelled from 5.63 years for RFS and 11.00 years for OS. The ERG believes that whilst it is possible that a proportion of patients might achieve cure following relapse (due to downstream HSCT), which may justify the use of models which imply different timepoints for cure (as suggested within the company's clarification response,[17] question B6), it is not clinically plausible to apply models which feature such a large gap between those achieving cure pre- and post-relapse.

Whilst the ERG recognises the difficulties of generating robust survival models given the evidence available, the ERG would have preferred the adoption of a model structure which aligns directly with the clinical input received by the company and the ERG - that patients who have not relapsed within 5-years are considered to be cured. However, the ERG notes that owing to the use of a partitioned survival model, applying the assumption of cure at 5-years to the OS curves produces a bias, as patients who are alive and relapsed at this timepoint gain additional survival benefit. The use of an alternative model structure (e.g. a state transition model) would rectify this problem, but may introduce alternative issues due to small sample sizes and an increased risk of selection bias.

Figure 13: RFS and OS cure points – blinatumomab group

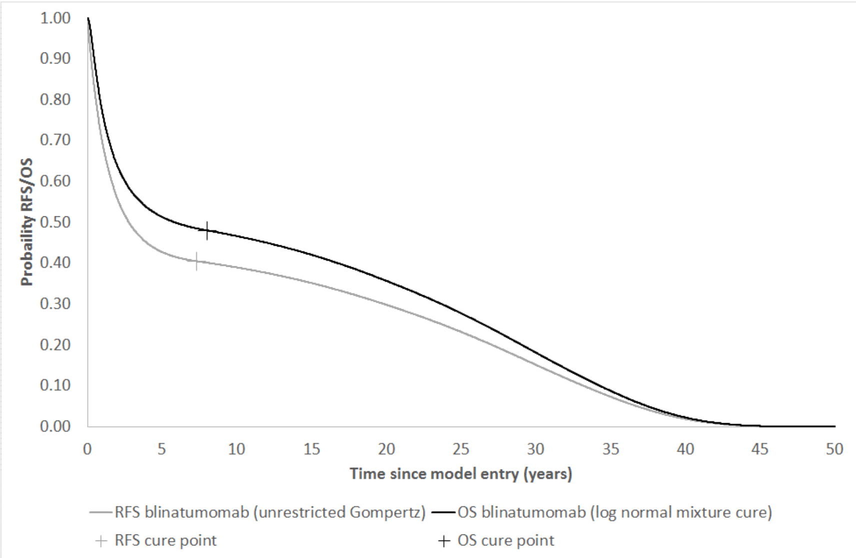
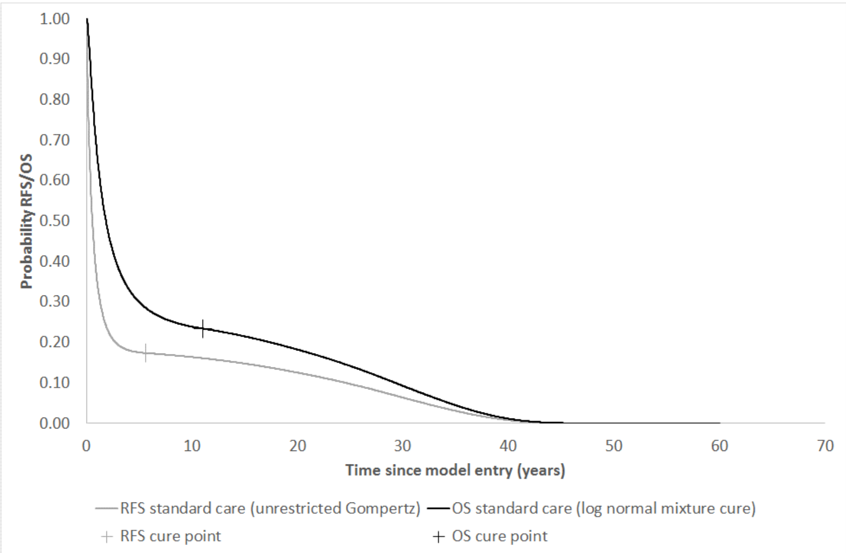


Figure 14: RFS and OS cure points – standard care group



On the basis of clinical advice received by the company and the ERG, the ERG considers that it would be more appropriate to apply a fixed cure point at 5-years.

(6) Concerns regarding the robustness of the company's alternative base case (blinatumomab used on relapse for the standard care group)

The CS[1] presents a "key scenario analysis" (which is referred to as an "alternative base case" in the company's clarification response,[17] question A2) in which blinatumomab is assumed to be used as first salvage therapy for 70% of patients who relapse on standard care chemotherapy. This analysis is based on the incremental survival gains, QALY gains and costs of blinatumomab (versus FLAG-IDA) from TA450;[53] these are added in to the base case total health gains and costs. This analysis produces an ICER for blinatumomab versus standard care of £17,420 per QALY gained (see Table 43).

Table 43: Company's alternative base case (blinatumomab used on relapse for standard care group)

Option	QALYs	Costs	Inc. QALYs	Inc. Costs	Incremental cost per QALY gained
Blinatumomab	7.10	[redacted]	1.91	£33,473	**£17,420**
Standard care	5.19	[redacted]	-	-	-

The ERG considers that this analysis is problematic for two reasons. Firstly, the MRD+ relapsed population within the current model reflects only a subgroup of the relapsed/refractory population within the TA450 model.[53] Secondly, the additional costs and health outcomes assumed to be related to blinatumomab salvage therapy are not structurally related to the OS gains estimated from the company's survival modelling of data from the BLAST PAS and the historical control study DCAS. As such, the ERG considers the results of this analysis to be highly uncertain.

(7) Questionable reliability of the company's HRQoL estimates

The ERG has several concerns regarding the plausibility of the HRQoL estimates assumed within the model.

(i) Relapse-free utility

The GLM/GEE-derived utility values for the RFS state (utility=0.79-0.84 depending on cycle, treatment and whether the patient has discontinued treatment) are similar to the general population utilities reported by Kind *et al*[50] (utility = 0.844). The clinical advisors to the ERG considered the relapse-free utility to be a reasonable reflection of the HRQoL for this population.

(ii) Post-relapse utility

The ERG has some concerns regarding the post-relapse utility value estimated using the logistic regression of the TOWER and BLAST studies.[1, 61] Clinical advisors to the ERG considered that the post-relapse utility estimate of 0.692 appears to be unrealistically high. The CS provides only limited details regarding the derivation of this estimate. During clarification (see clarification response,[17] question B20), the ERG requested further information regarding the observed EQ-5D estimates in the relapsed population of BLAST. In response, the company stated there were only 8 post-relapse utility assessments, of which 6 assessments were conducted on the day of relapse, 1 assessment was conducted 22 days after relapse, and 1 assessment was conducted 30 days after relapse. The mean utility value for these 8 post-relapse assessments was 0.819 (0.276). The ERG notes that this value is higher than some of the relapse-free utility estimates derived from the company's GLM/GEE model and that this estimate is therefore not reliable. The ERG's clinical advisors suggested that HRQoL in relapsed patients would likely be much lower (estimated utility=0.25 to 0.60), irrespective of whether the patient was fit enough for transplant.

(iii) General population HRQoL

The company's original base case model used the study reported by Kind *et al*[50] to estimate age- and sex-specific general population health utilities. The ERG considers the regression study of Health Survey for England (HSE) data reported by Ara and Brazier[58] to represent a more appropriate source for these parameters, as it includes a larger sample size (Ara and Brazier n=26,729; Kind *et al* n=3,395) and it is more up-to-date (Ara and Brazier, 2010 [based on HSE 2003 and 2006]; Kind *et al* 1999 [based on data collected in 1993]). In response to a request for clarification[17] (question B21), the company updated their model to use HRQoL estimates from Ara and Brazier.[58] The company's clarification response[17] notes that the utility values from Ara and Brazier[58] are generally slightly higher than those based on Kind *et al;*[50] as such, the use of these newer estimates yields a slightly more favourable ICER for blinatumomab versus standard care compared with the company's original base case (ICER using Ara and Brazier=£27,938 per QALY gained; ICER using Kind *et al*=£28,524).

(iv) Decrement associated with exposure to radiotherapy, chemotherapy, and HSCT

The ERG considers that the HRQoL decrement associated with radiotherapy, chemotherapy, and HSCT, which is based on a mid-point value, is essentially arbitrary. The ERG also notes that during the first 5-years post-HSCT, the proportion of this decrement which is attributable to HSCT should already be captured through the QALY losses estimated through the HSCT sub-models.

(8) Uncertainty surrounding the proportion of RFS events that are deaths

The ERG notes that there is a considerable difference in the proportion of RFS death events between the data from BLAST and the ATT-weighted data from the historical control study DCAS (BLAST PAS RFS death probability = 47.1%, historical control DCAS RFS death probability = 8.5%). The CS makes the case that the high probability observed in BLAST may be a consequence of incomplete capture of relapses after transplant in BLAST and mismatched donors resulting in infections.

The ERG agrees that there may be issues surrounding incomplete data collection in BLAST, as the level of censoring is considerably higher than in the historical control study DCAS (BLAST PAS total n=34 events; ATT-weighted historical control study total n=122.3 events). However, it is not clear that the proportion of death events would necessarily decrease with additional follow-up. Furthermore, it is unclear from the CS whether infections caused by mismatched donors and intensive immunosuppression were the cause of death in these patients. The ERG notes that decreasing the RFS death proportion in the blinatumomab group leads to a less favourable ICER.

(9) Unrealistic treatment pathway

The company's model captures a single treatment pathway for the standard care comparator. This is assumed to be comprised of chemotherapy according to the UKALL14 maintenance therapy regimen[19] (vincristine, methotrexate [intrathecal], prednisolone, mercaptopurine and methotrexate [oral]) followed either by HSCT(s) and/or salvage chemotherapy (FLAG-IDA). Clinical advice received by the ERG suggests that the treatment pathways for patients with Ph- MRD+ BCP-ALL are more complex and depend on the patient's level of MRD positivity, patient fitness, their eligibility for allograft (including the availability of matched donors), as well as variability between centres and paediatric and adult haematologists.

Clinical advisors to the ERG provided the following description of the treatment pathway for patients with MRD+ BCP-ALL.

At present, adults aged 16-60 years being treated for ALL with curative intent in the UK will receive intensive chemotherapy that can broadly be described in 4 phases – induction, intensification and consolidation followed by maintenance. Different terms are used in the paediatric protocol although the chemotherapies used are similar. Although there is no routine allografting in the paediatric protocol UKALL2011, the current adult protocol UKALL14 stipulates that most adults receive an allogeneic transplant rather than continue with chemotherapy alone. Allografting usually occurs post intensification in place of consolidation and maintenance.

Patients that have persistent MRD following induction chemotherapy are at an increased risk of relapse and will usually require an allogeneic transplant to have any chance of cure. The exception to this is the younger teenage patients with low levels of MRD $<10^{-3}$ where it may be acceptable to continue with chemotherapy only in some circumstances. The success of allografting in adults is directly linked to the levels of MRD prior to the transplant. Adult patients with persistent MRD $<10^{-3}$ may be cured by an allograft (although this chance is increased if MRD can be reduced to $<10^{-4}$). Those under 40 years of age would be suitable to go straight to a myeloablative transplant at this stage. Those over 40 years of age would have intensification with high-dose methotrexate as an inpatient before receiving a reduced intensity allogeneic transplant.

Transplantation is unlikely to be curative when the levels of MRD post induction are 10^{-3} or higher. In this situation, patients will require more intensive blocks of salvage chemotherapy as an inpatient in order to try and reduce the levels of MRD prior to allografting. However, these patients may have chemo-refractory disease and may not be able to achieve deeper levels of MRD in which case an early relapse is likely.

Those patients that have persistent MRD and are not able to proceed to an allograft for any reason e.g. no suitable donor, failure to reduce MRD to an acceptable level or poor general fitness, will be given standard chemotherapy in an attempt to prolong life although this strategy is unlikely to be curative.

The ERG's clinical advisors also noted that patients would receive FLAG-IDA as salvage chemotherapy; after failing this regimen, a different regimen would be used.

The ERG therefore has concerns that the company's model does not fully reflect the complexity of current treatment pathways followed by patients in England. Specifically, the ERG notes the following:

- The company's model only includes a single standard care chemotherapy regimen
- The model does not reflect any interplay between patient characteristics (e.g. age, fitness, eligibility for HSCT) and treatments received.
- The company's assumption that patients who fail FLAG-IDA salvage would receive further therapy using this regimen is inappropriate.

(10) Limited sensitivity analysis around alternative parametric functions

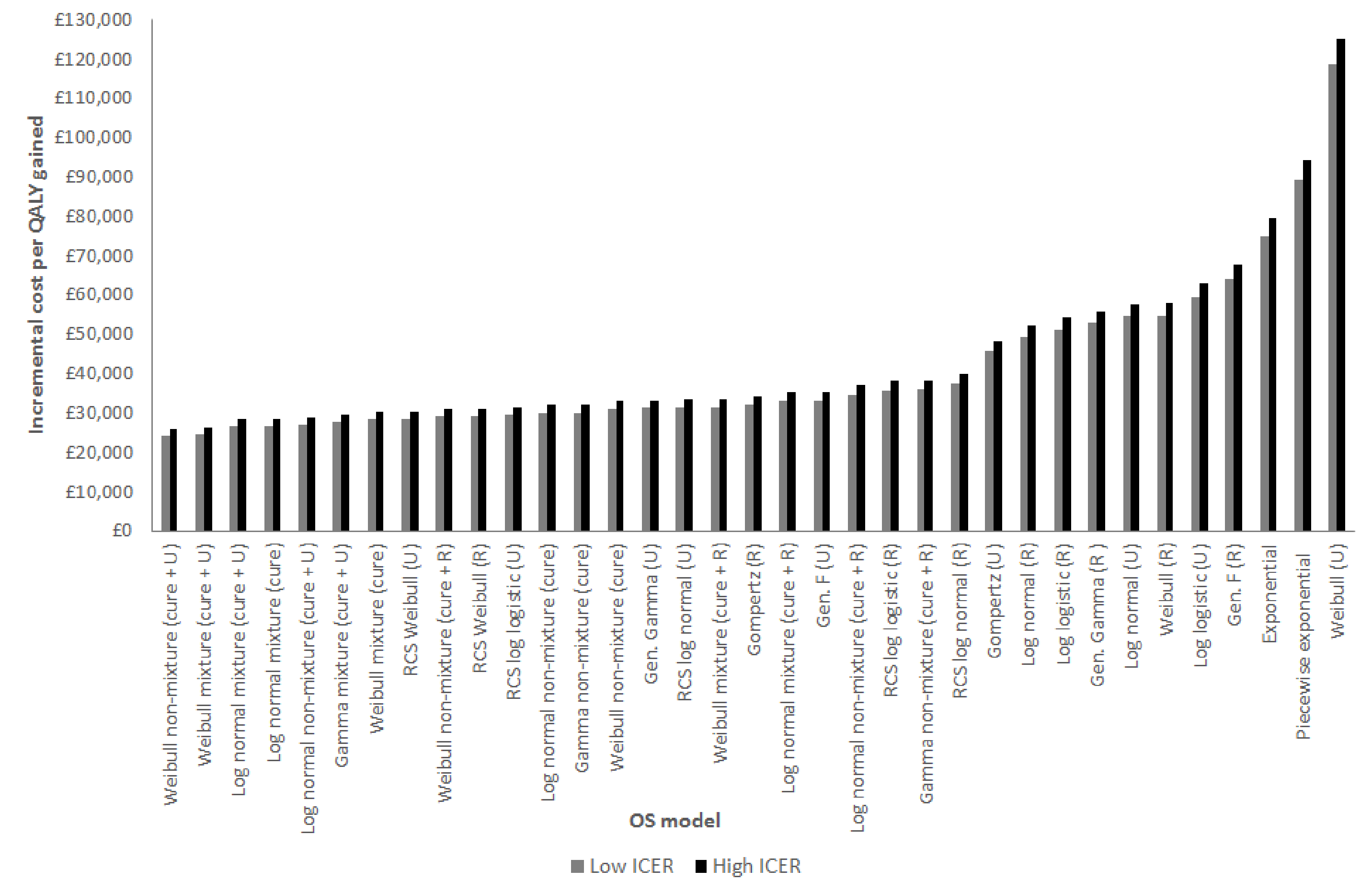
The company fitted 38 separate models to RFS and 35 separate models to OS. Whilst this indicates that there are many possible combinations of potentially plausible RFS and OS models, the CS includes only two additional scenario analyses which explore the impact of using alternative parametric functions for RFS and OS:

(i) RFS and OS distributions changed to restricted Gompertz and unrestricted Weibull non-mixture cure, respectively. The ICER for this scenario is reported to be £25,081 per QALY gained.

(ii) RFS and OS distributions changed to restricted RCS log-logistic and restricted RCS Weibull, respectively. The ICER for this scenario is reported to be £30,647 per QALY gained.

In response to a request for clarification[17] (question B8), the company presented analyses which combine different RFS and OS models across 1,330 different combinations. Figure 15 presents the distribution of the resulting ICERs according to the selected OS function, with low and high ICERs indicating the impact of assuming different RFS functions given the selected OS function. The Weibull non-mixture cure (unrestricted) OS model produces the lowest ICER (£24,171 per QALY gained); the Weibull (unrestricted) OS model produces the highest ICER of (£125,153 per QALY gained). The highest ICER arising from any OS cure model is £38,076 per QALY gained. The ERG notes that the company's base case ICER is towards the lower end of the range of possible ICERs.

Figure 15: ICERs by alternative parametric OS model from company's clarification response (range determined by RFS curve selected)



5.4 ERG exploratory analyses

ERG exploratory analyses - methods

The ERG undertook eight sets of exploratory analysis. All analyses were undertaken using the deterministic version of the updated model submitted by the company following clarification.[17] As the bootstrap RFS and OS samples for the company's base case model selections were hardcoded into the model, it was not possible to re-run the probabilistic model using alternative RFS/OS functions. Technical details relating to the implementation of these analyses can be found in Appendix 1.

The ERG's analyses include two key exploratory analyses which when combined represent the ERG-preferred model. These analyses are detailed below:

Exploratory analysis 1: Correction of errors. Within this analysis, seven programming errors identified during the ERG's double-programming and model verification exercise were rectified (see Section 5.3.3).

Exploratory analysis 2: Inclusion of a fixed cure point of 5-years. Within this analysis, the hazard of death for all patients surviving beyond 5-years was switched from the hazard predicted by the parametric model to that of the uplifted general population (where a given model has a cure point which manifests at less than 5 years from model entry, this assumption has no effect). This amendment was implemented using existing functionality contained within the company's model. It should be noted that this assumption better reflects clinical judgement and means that the cure point is applied structurally at a fixed timepoint, rather than being determined by the statistical model. However, due to limitations in the company's model structure, the ERG does not consider this analysis to be ideal, as the cure is applied to both RFS and OS functions at the cure point; patients who are alive and have relapsed at 5-years (9% of the blinatumomab group and 13% of the standard care group) will therefore be considered cured, which is not realistic. Given the model structure, it was not possible to relax this assumption. Therefore, this analysis will produce a bias in favour of the standard care group, although the ERG considers the magnitude of this is likely to be small.

Exploratory analysis 3: ERG-preferred model. This analysis combines exploratory analyses 1 and 2. Notwithstanding the uncertainty surrounding the selection of parametric RFS and OS functions, this analysis represents the ERG's preferred model.

In addition, five further sets of exploratory analyses were undertaken using the ERG's preferred model:

Exploratory analysis 4: Exploration of impact of alternative standard care chemotherapy costs. Within this analysis, the drug acquisition costs for standard care chemotherapy were doubled in order to assess the impact of assuming alternative treatment regimens on the ICER for blinatumomab versus standard care.

Exploratory analysis 5: Exploration of the impact of alternative post-HSCT survival probabilities. Within this analysis, post-HSCT survival was estimated using data on the 100-day mortality rate after allogenic HSCT from BLAST[39] and the uplifted age- and sex-weighted general population mortality rates thereafter. For the first 6-monthly cycle post-HSCT, the probability of death was calculated by adding the 100-day mortality rate from BLAST to the probability of death in the remaining 82.6 days of the cycle using the uplifted general population mortality rates. For all subsequent 6-monthly cycles, the probability of death was estimated using the uplifted general population mortality rates.

Exploratory analysis 6: Exploration of alternative cure fractions for the standard care group. This analysis was undertaken to assess the sensitivity of the model results to the assumed cure fraction for the standard care group. Analyses were undertaken for cure fractions of 25%, 30% and 35%.

Exploratory analysis 7: Exploration of alternative post-relapse HRQoL estimates. Within this analysis, three alternative HRQoL estimates were applied to the post-relapse state in order to explore their impact on the ICER for blinatumomab versus standard care: (i) the observed EQ-5D value for the small number of patients with post-relapse utility assessments in BLAST;[1] (ii) an assumed value of 0.50 and (iii) an assumed value of 0.25.

Exploratory analysis 8: Exploration of the impact of alternative parametric RFS and OS models. Within this analysis, the model was run assuming alternative unrestricted parametric OS and RFS models across a total of 1,330 model combinations. Clinical advisors to the ERG were asked to select their preferred unrestricted OS function and to give reasons supporting their selections (see Figure 16 and Figure 17 for survival plots; full model selection questionnaire presented in Appendix 2).

Figure 16: Predicted cumulative survival probabilities by OS model type (including 5-year cure assumption and mortality risk in cured population) - blinatumomab

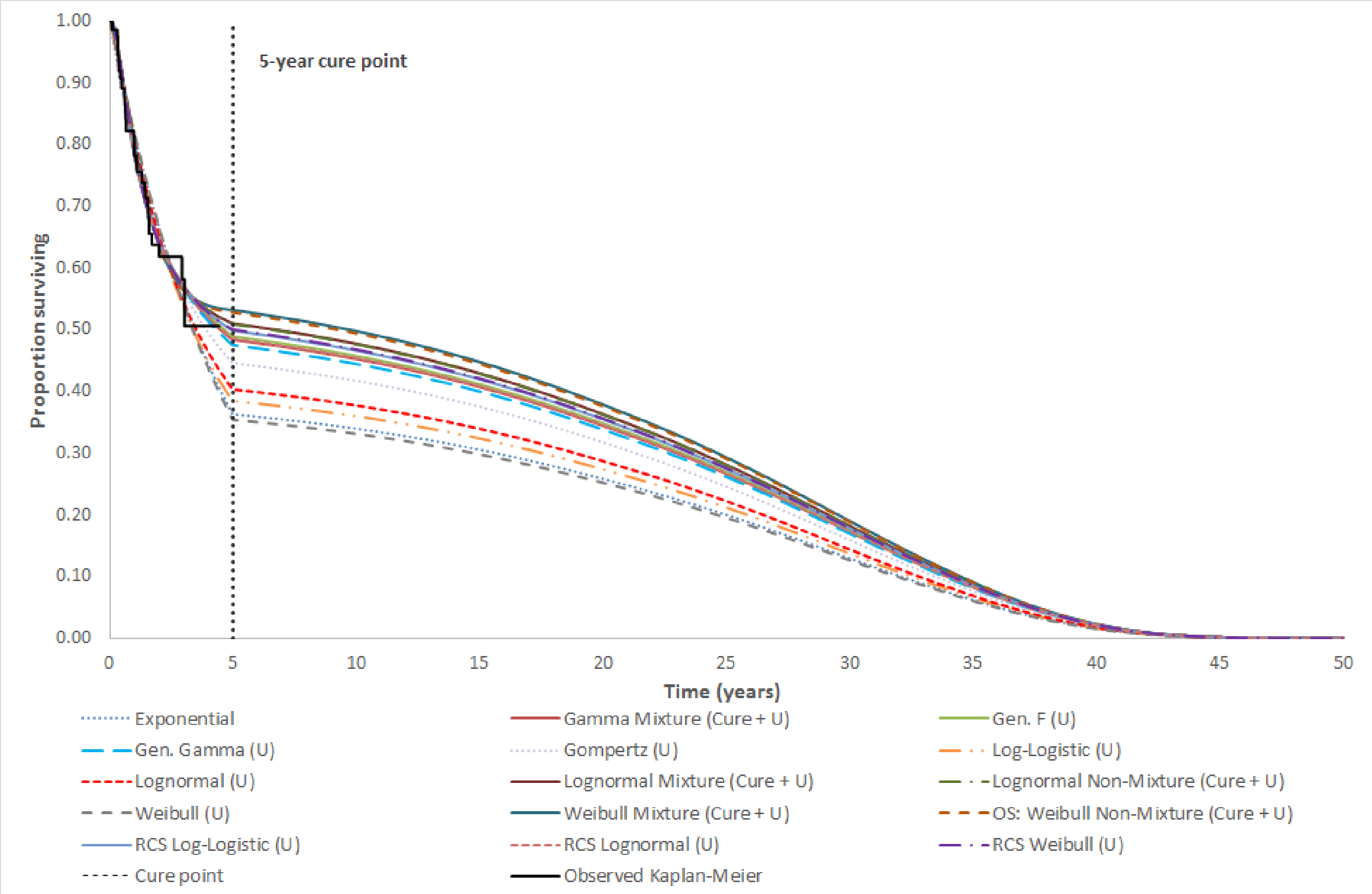
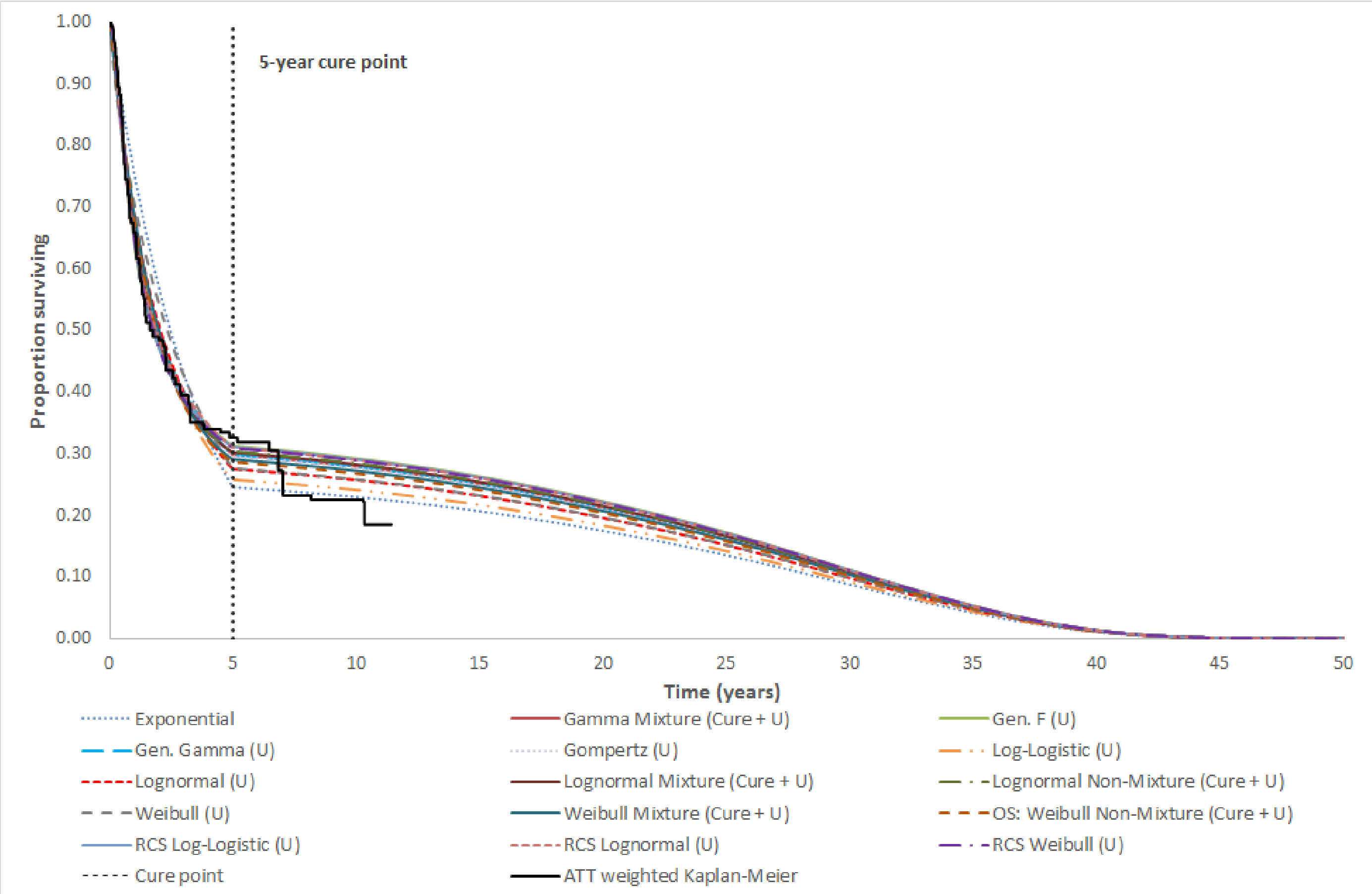


Figure 17: Predicted cumulative survival probabilities by OS model type (including 5-year cure assumption and mortality risk in cured population) – standard care



ERG exploratory analyses - results

ERG exploratory analyses 1-3 – Correction of errors and inclusion of a 5-year cure point

Table 44 presents the results of exploratory analyses 1-3. Analyses 1 and 2 are applied individually to the company's updated model submitted post-clarification; analysis 3 combines both analyses to reflect the ERG's preferred model. As shown in Table 44, the correction of errors has only a minor impact upon the ICER for blinatumomab versus standard care (ICER=£27,717 per QALY gained). The incorporation of an assumption of cure at 5-years also leads to a slightly less favourable ICER for blinatumomab versus standard care (ICER=£30,304 per QALY gained). When these analyses are combined, the deterministic ICER for blinatumomab versus standard care is estimated to be £30,227 per QALY gained.

Table 44: Results of ERG exploratory analyses 1-3 (error correction and inclusion of a 5-year cure point)

Option	QALYs	Costs	Inc. QALYs	Inc. Costs	Incremental cost per QALY gained
Company's base case (updated model)					
Blinatumomab	7.23	[redacted]	3.02	£83,800	**£27,779**
Standard care	4.21	[redacted]	-	-	-
ERG exploratory analysis 1 – Correction of errors identified during model verification					
Blinatumomab	7.21	[redacted]	3.00	£83,264	**£27,717**
Standard care	4.21	[redacted]	-	-	-
ERG exploratory analysis 2 – Cure applied to all surviving patients at 5 years					
Blinatumomab	7.37	[redacted]	2.77	£83,803	**£30,304**
Standard care	4.61	[redacted]	-	-	-
ERG exploratory analysis 3 –Analyses 1 and 2 combined (ERG-preferred model)					
Blinatumomab	7.35	[redacted]	2.75	£83,268	**£30,227**
Standard care	4.59	[redacted]	-	-	-

Further sensitivity analyses undertaken using the ERG-preferred model

Table 45, Table 46, Table 47, Table 48, Table 48 and Figure 18 present additional sensitivity analyses around the ERG's preferred model in order to explore the impact of alternative assumptions of the ICER for blinatumomab versus standard care.

ERG exploratory analysis 4 – Standard care chemotherapy costs doubled

Table 45 presents the results of an analysis in which the costs of standard care chemotherapy were doubled. This analysis suggests that the costs of standard care chemotherapy do not materially impact upon the ICER for blinatumomab.

Table 45: ERG exploratory analysis 4 – Standard care chemotherapy costs doubled (based on the ERG-preferred model)

Option	QALYs	Costs	Inc. QALYs	Inc. Costs	Incremental cost per QALY gained
Blinatumomab	7.35	[redacted]	2.75	£82,222	**£29,848**
Standard care	4.59	[redacted]	-	-	-

ERG exploratory analysis 5 – Use of alternative HSCT survival probabilities

Table 46 presents the results of an analysis in which the probability of remaining alive and in follow-up following HSCT were increased, based on the 100-day mortality rate for blinatumomab and uplifted general population mortality rates. This analysis indicates that the HSCT survival probabilities lead to an increase in the ICER for blinatumomab versus standard care, although the ERG notes there is uncertainty surrounding the survival trajectory of patients undergoing HSCT.

Table 46: ERG exploratory analysis 5 – Use of alternative HSCT survival probabilities (based on the ERG-preferred model)

Option	QALYs	Costs	Inc. QALYs	Inc. Costs	Incremental cost per QALY gained
Blinatumomab	7.29	[redacted]	2.73	£89,302	**£32,667**
Standard care	4.55	[redacted]	-	-	-

ERG exploratory analysis 6 – Use of alternative cure fractions for standard care chemotherapy

Table 47 presents the results of the analyses whereby the cure fraction for the standard care group was set equal to 0.25, 0.30 and 0.35, respectively. The results of the analysis highlight that the cure fraction is a key driver of cost-effectiveness for blinatumomab versus standard care.

Table 47: ERG exploratory analysis 6 – Use of alternative cure fractions for standard care chemotherapy (based on the ERG-preferred model)

Standard care cure fraction	Blinatumomab versus standard care		
	Inc. QALYs	**Inc. costs**	**ICER**
Cure fraction = 0.21 (company's base care)	2.75	£83,268	**£30,227**
Cure fraction = 0.25	2.36	£81,402	**£34,465**
Cure fraction = 0.30	1.83	£78,883	**£43,072**
Cure fraction = 0.35	1.30	£76,363	**£58,697**

Exploratory analysis 7 – Impact of alternative post-relapse utility values

Table 48 presents the results of the analyses in which alternative post-relapse utility values are applied. As shown in the table, the post-relapse utility value has a fairly minor impact on the ICER, with lower values resulting in more favourable ICERs for blinatumomab versus standard care. The ERG notes that even at extreme values of post-relapse utility (for example, utility=0.25), the ICER is reduced only by around £3,000.

Table 48: Exploratory analysis 7 – Impact of alternative post-relapse utility values (based on the ERG-preferred model)

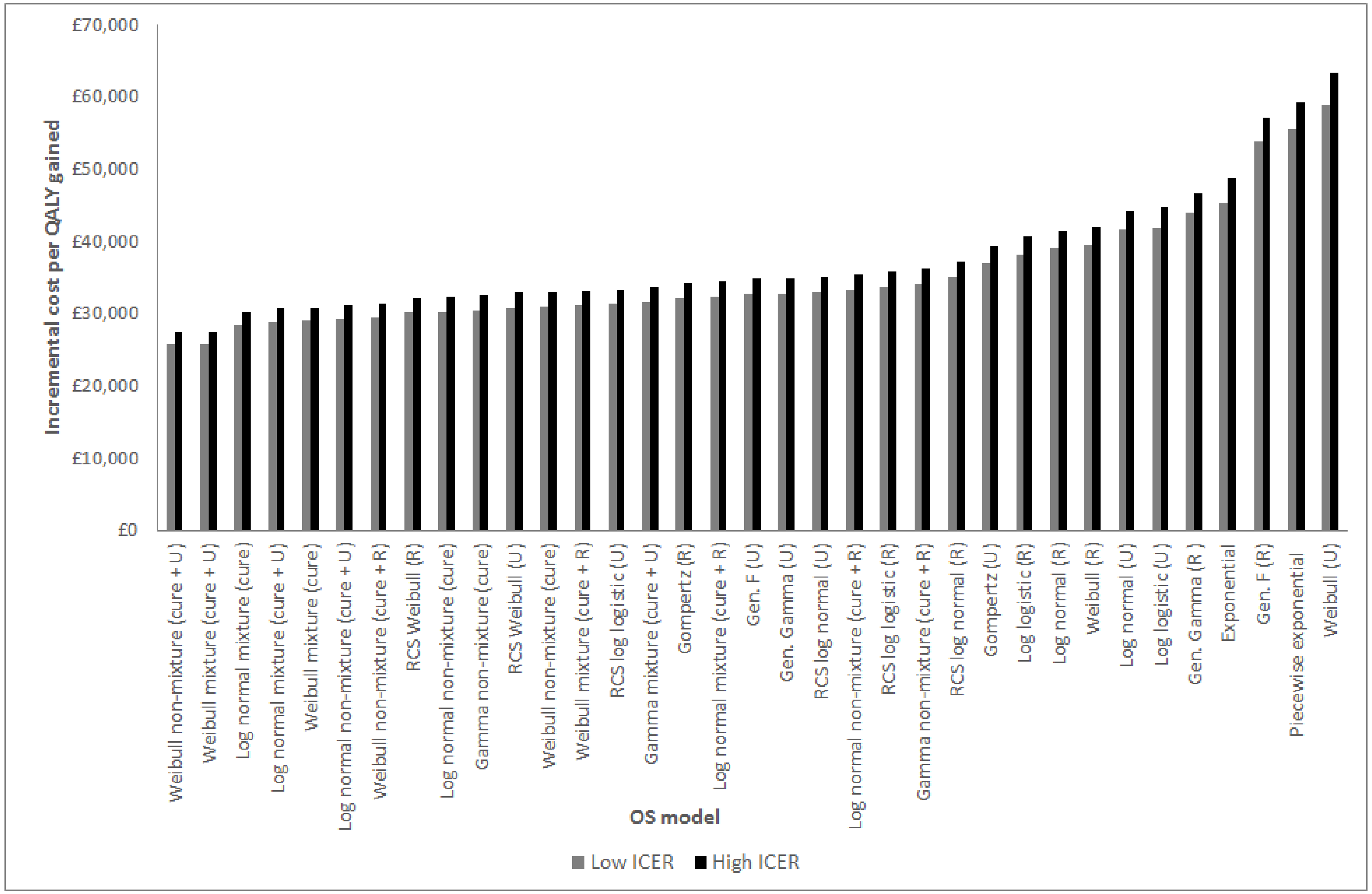
Post-relapse utility value	Blinatumomab versus standard care		
	Inc. QALYs	**Inc. costs**	**ICER**
Utility=0.69 (company's base case)	2.75	£83,268	**£30,227**
Utility=0.819 (BLAST post-relapse utility[1])	2.67	£83,268	**£31,157**
Utility=0.50	2.88	£83,268	**£28,930**
Utility=0.25	3.04	£83,268	**£27,395**

ERG exploratory analysis 8 - Impact of alternative parametric RFS and OS models on the ICER for blinatumomab

Figure 18 presents the results of additional analyses of the ERG's preferred model in which a large range of alternative parametric models are assumed for RFS and OS. For each OS model, the range of low and high ICERs reflects the impact of assuming alternative RFS functions. Based on the ERG's preferred model, this exploratory analysis indicates the following:

- The inclusion of the 5-year cure assumption reduces the variation in ICERs across the OS models considered. The ICER for blinatumomab versus standard care ranges from £25,783 per QALY gained (Weibull non-mixture cure model, unrestricted) to £63,265 per QALY gained (Weibull, unrestricted).
- As with the company's analyses presented in Section 5.3, for a given OS model, the RFS function does not generally produce a large range in terms of the highest and lowest ICER. The ICER range for RFS given the selected OS model is typically around £2,000.
- In general, the cure models produce lower ICERs than the other OS functional forms (standard parametric models and RCS models).
- Only the Weibull non-mixture cure model (unrestricted) and the Weibull mixture cure model (unrestricted) produce results in which the full range of ICERs are below £30,000 per QALY gained.
- The ICERs at the lower end of the range for the log normal mixture (cure), log normal mixture (cure, unrestricted), Weibull mixture (cure) and log normal non-mixture (cure, unrestricted) and Weibull non-mixture (cure, unrestricted) are below £30,000 per QALY gained.

Figure 18: ERG exploratory analysis 8 - Impact of alternative parametric RFS and OS models on the ICER for blinatumomab (low-high ICER range determined by RFS curve given the selected OS model)



The clinical advisors to the ERG considered the assumption of a cure point at 5 years to be acceptable and noted that this is in line with data observed in the UKALLXII trial,[62] whereby the Kaplan-Meier OS curves begin to approach an approximate plateau from around year 3.

Advice on the range of plausible statistical models was also consistent. For blinatumomab, the preferred distribution given by one advisor was the generalised gamma. This was selected on the basis that the estimated OS of 50% at 5 years was considered to concord with the observed data from BLAST and study MT103-202. The ERG's second clinical advisor selected the RCS Weibull model and the log normal mixture/non-mixture cure models as their preferred choice based on these providing clinically expected changes in OS between years four and five. It was noted that after four years, the rate of events would be expected to decrease, in line with a cure point at around five years. Models suggesting a steep drop in OS during this interval were considered implausible as these provide predictions with an unrealistic change in the hazard rate at 5 years when combined with the elevated general population mortality estimates. This led both clinical advisors to dismiss the four lowest predicting models (log normal, log logistic, exponential and Weibull). The first advisor also stated that models which provided a more favourable OS profile than the RCS Weibull (Weibull mixture cure [unrestricted], Weibull non-mixture cure [unrestricted], log normal mixture cure [unrestricted], log normal non-mixture cure [unrestricted]) were unlikely to be plausible.

For standard care chemotherapy, the Weibull mixture cure was selected as the preferred distribution by the first clinical advisor, based on its predicted 5-year OS probability. Models between the log normal and RSC Weibull were considered to be plausible. The second clinical advisor selected the RSC Weibull as the preferred distribution based on the fit to the (ATT-weighted) observed data up to five years. Both clinical advisors expressed uncertainty in the clinical plausibility of the observed drop in OS from year 6 onwards, which did not reflect their experience in clinical practice by which time very few events would be expected.

Table 49: ERG clinical advisors' list of potentially plausible OS models (preferred model highlighted in bold)

OS model	Clinical advisor 1	Clinical advisor 2
Blinatumomab	RCS Weibull (U) RCS log logistic (U) Generalised F (U) RCS log normal (U) Gamma mixture cure (U) **Generalised Gamma (U)** Gompertz (U)	Log normal mixture cure (U) Log normal non-mixture cure (U) **RCS Weibull (U)** RCS log logistic (U) Generalised F (U) RCS log normal (U) Gamma mixture cure (U) Generalised gamma (U) Gompertz (U)
Standard care	RCS Weibull (U) Log normal non-mixture cure (U) Log normal mixture cure (U) Generalised gamma (U) Gamma mixture cure (U) Gompertz (U) **Weibull mixture cure (U)** OS: Weibull non-mixture cure (U) Weibull (U) Log normal (U)	**RCS Weibull (U)** No clear range given due to similarity of curves

U – unrestricted; RCS – restricted cubic spline

Table 50 summarises the ICER ranges associated with the three OS models preferred by the ERG's clinical advisors. The clinical advisors' three preferred OS models (Generalised gamma [unrestricted], RCS Weibull [unrestricted] and Weibull mixture cure [unrestricted]) result in ICERs in the range £25,810 per QALY gained to £34,904 per QALY gained.

Table 50: ICERs associated with clinical advisors' preferred OS functions (low-high ICER range determined by RFS curve given the selected OS model)

OS model	Low ICER	High ICER
Generalised gamma (U)	£32,800	£34,904
RCS Weibull (U)	£30,868	£32,857
Weibull Mixture (Cure + U)	£25,810	£27,492

U – unrestricted

5.5 Discussion

The CS includes a systematic review of published economic evaluations of treatments for adult ALL patients with MRD-positivity after treatment together with a *de novo* health economic analysis of blinatumomab versus standard care chemotherapy in patients with Ph- MRD+ BCP-ALL. The company's review did not identify any published economic evaluations of blinatumomab in this indication.

The company's *de novo* partitioned survival model assesses the cost-effectiveness of blinatumomab versus chemotherapy (based on the UKALL14 maintenance regimen) in patients with Ph- MRD+ BCP-

ALL in CR1. Incremental health gains, costs and cost-effectiveness of blinatumomab are evaluated over a 50-year time horizon from the perspective of the NHS and PSS. The company's model is comprised of a main structure which reflects RFS and OS outcomes, as well as two linked sub-models which are intended to estimate additional costs and HRQoL decrements associated with HSCT given before and/or after relapse. The main model structure includes three health states: (1) relapse-free; (2) post-relapsed and (3) dead. The model parameters were informed by analyses of time-to-event data (RFS and OS) from the company's IPTW weighted analysis of the BLAST PAS and the ATT-weights historical control study DCAS. RFS is modelled using an unrestricted Gompertz distribution (using an approach which is analogous to fitting models independently to each treatment group), whilst OS is modelled using a log normal mixture cure model (whereby the treatment effect is applied only to the cure fraction parameter). HRQoL is assumed to be principally determined by relapse status, time spent in the relapse-free state and treatment received; utility estimates were derived from a GLM/GEE model fitted to EQ-5D data collected in BLAST, a propensity matching analysis of the BLAST and TOWER blinatumomab studies, as well as other literature and assumptions. Resource use estimates and costs were based on data collected in BLAST, the UK ALL14 treatment protocol, routine cost sources, clinical opinion and other literature.

Based on the probabilistic version of the model (assuming the unrestricted Gompertz function for RFS and the log normal mixture cure model for OS), blinatumomab is expected to generate an additional 2.85 QALYs at an additional cost of £84,456 compared with standard care: the corresponding ICER for blinatumomab versus standard care is £29,673 per QALY gained. The deterministic version of the company's model produces a similar ICER of £28,524 per QALY gained for blinatumomab versus standard care. Assuming a WTP threshold (λ) of £20,000 per QALY gained, the company's model suggests that the probability that blinatumomab produces more net benefit than standard care is 0.10. Assuming a WTP threshold of £30,000 per QALY gained, the probability that blinatumomab produces more net benefit than standard care is estimated to be 0.53. Following the clarification process, the company submitted a revised model which addressed some of the minor concerns initially raised by the ERG; this updated model generated an ICER for blinatumomab versus standard care of £28,655 per QALY gained.

The ERG critically appraised the company's economic analysis and double-programmed the deterministic version of the company's model. The ERG's critical appraisal identified several issues relating to the company's economic analysis and the evidence used to inform it. These include: (i) the exclusion of relevant patient subgroups from the model; (ii) the exclusion of the "monitor for relapse" comparator from the analysis; (iii) use of a model structure which is inappropriate for tracking HSCT; (iv) the absence of RCT evidence for blinatumomab versus standard care; (v) concerns regarding the company's approach to RFS/OS model selection; (vi) concerns regarding the robustness of the company's alternative base case (blinatumomab used on relapse for the standard care group); (vii) the questionable reliability of the company's HRQoL estimates; (viii) uncertainty surrounding the proportion of RFS events that are deaths; (ix) the inclusion of an unrealistic treatment pathway and (x) limited sensitivity analysis around alternative parametric functions.

The ERG undertook eight sets of exploratory analyses using the deterministic version of the company's updated model. Notwithstanding uncertainty relating to the choice of parametric RFS and OS functions, the ERG's preferred model includes the correction of seven minor programming errors and the inclusion of a fixed 5-year cure point. The ERG-preferred model produces a deterministic ICER for blinatumomab versus standard care of £30,227 per QALY gained. The ERG also undertook a number of further analyses to explore the impact of alternative parametric models and alternative parameter values on the model results. These analyses indicate that the costs of standard chemotherapy, the post-HSCT survival probabilities and the utility value for the post-relapse state have only a minor impact on the ICER for blinatumomab versus standard care. Conversely, the cure fraction and the choice of parametric OS distribution have a significant impact on the ICER for blinatumomab versus standard care. Within the ERG's exploratory analysis of alternative RFS and OS functions, the ICER for blinatumomab versus standard care ranges from £25,783 per QALY gained (Weibull non-mixture cure model, unrestricted) to £63,265 per QALY gained (Weibull, unrestricted). Across the full range of models considered, only the Weibull non-mixture cure model (unrestricted) and the Weibull mixture cure model (unrestricted) produce results in which the full range of ICERs are below £30,000 per QALY gained (irrespective of RFS model assumed). The clinical advisors' three preferred OS models (Generalised gamma [unrestricted], RCS Weibull [unrestricted] and Weibull mixture cure [unrestricted]) result in ICERs in the range £25,810 per QALY gained to £34,904 per QALY gained.

The ERG considers the following to represent the key uncertainties within the company's health economic analysis:

- The absence of comparative clinical and economic evidence for blinatumomab versus standard care chemotherapy within subgroups of BLAST which were excluded from the comparative analysis (patients with Ph+ MRD+ BCP-ALL and patients with Ph- MRD+ BCP-ALL with CR2+).
- The absence of clinical data and economic comparisons of blinatumomab versus monitoring for patients who are unable to undergo HSCT or to tolerate chemotherapy.
- The necessary reliance on adjusted historical control evidence, due to the absence of RCT evidence for blinatumomab versus standard care, and the potential for unobserved confounders.
- The long-term extrapolation of RFS and OS outcomes, including the timing of cure.

6. END OF LIFE

NICE end of life supplementary advice should be applied in the following circumstances and when both the criteria referred to below are satisfied:

- The treatment is indicated for patients with a short life expectancy, normally less than 24 months and;
- There is sufficient evidence to indicate that the treatment offers an extension to life, normally of at least an additional 3 months, compared to current NHS treatment.

The CS[1] states that blinatumomab meets NICE's criteria for life-extending therapies given at the end of life. The company's evidence supporting this is presented in Table 51.

Table 51: Evidence supporting the company's end of life argument (reproduced from CS Table 50)

Criterion	Data available
The treatment is indicated for patients with a short life expectancy, normally less than 24 months	Median OS, using ATT-weighted propensity score matching analyses for standard care chemotherapy was [redacted]. The estimated mean survival (undiscounted) in the economic analysis was almost 5x greater than the median survival (7.86 years) in the standard care arm; however, this is reflective of the small proportion of patients who achieve long-term survival (~20%). For this reason, the median survival is considered to be a more suitable representation of the anticipated survival in the patient population as a whole.
There is sufficient evidence to indicate that the treatment offers an extension to life, normally of at least an additional 3 months, compared with current NHS treatment	Median OS, using ATT-weighted propensity score matching analyses (Section B.2.9.4), was [redacted] after more than 40 months follow-up for blinatumomab thus demonstrating a [redacted] OS survival [redacted] when compared to standard care. The estimated mean survival (undiscounted) in the economic analysis was [redacted] years in the blinatumomab arm, resulting in an incremental survival benefit of [redacted] years.

The CS argues that a small number of patients in the historical control study who received standard care chemotherapy were observed to survive for a long time and that *"[Due to] the skew caused by this small group of patients, it was considered appropriate to use median OS values, rather than the mean, so as to more accurately represent the patient population as a whole. This skew effect and use of median OS rather than the mean has been noted in previous appraisals where the Committee agreed that consideration of medians was more appropriate"* (CS,[1] page 84).

The ERG strongly disagrees with the company's proposed use of median values to determine whether NICE's end of life criteria are met. Medians represent the "middle patient" and do not take account of skewness in the distribution of patient outcomes; conversely, the only measure of central tendency which fully represents outcomes for the population as a whole is the mean. Given the use of parametric cure models to inform OS, the mean and median OS estimates generated by the company's model diverge significantly (blinatumomab median OS=5.85 years versus mean [undiscounted] OS=13.59 years; standard care median OS=1.86 versus mean [undiscounted] OS=7.86 years). Based on the ERG's exploratory analyses, the lowest (undiscounted) mean OS for the standard care group across all models considered is 7.69 life years; all OS models suggest an undiscounted incremental OS gain of 2.12 years or greater. On the basis of these exploratory analyses, the ERG does not believe that blinatumomab meets NICE's criteria for life-extending treatments given at the end of life. The ERG also notes that due to the absence of a head-to-head RCT comparing blinatumomab against a relevant comparator, and the necessary use of a statistical matching approach to inform indirect treatment comparisons, there is uncertainty surrounding the true magnitude of OS benefit attributable to blinatumomab.

7. OVERALL CONCLUSIONS

In the absence of direct comparative data with other treatments, the main evidence in the CS was derived from two single-arm open-label studies of blinatumomab, of which one was a pilot study which was not used for the comparison with standard care chemotherapy. The two blinatumomab studies were well conducted, however single-arm studies are subject to biases. The main blinatumomab evidence came from the BLAST study of 116 patients. One historical control study (Study 20120148) of standard care chemotherapy was included (n=287); this study that analysed data from existing clinical databases.

From the 116 patients in BLAST, median OS was [redacted], with an OS at 18 months follow-up of [redacted]. From 110 patients providing RFS data from BLAST, median RFS was [redacted]; RFS at 18 months was [redacted]. Based on the EORTC QLQ-C30, patients reported [redacted]. By the end of the core study, [redacted] HRQoL as measured by EQ-5D did not change significantly from baseline to the end of the core study. [redacted] participants experienced at least one treatment-emergent AE. Comparative effectiveness for patients with Ph- MRD+ BCP-ALL in CR1 was estimated through indirect comparison of the BLAST PAS data and a historical control study using ATT propensity score weights. This analysis suggested an HR [redacted].

Notwithstanding uncertainty relating to the choice of parametric RFS and OS functions, the ERG's preferred analysis increases the ICER for blinatumomab versus standard care from £27,779 to £30,227 per QALY gained; this difference is driven by the inclusion of a structural cure assumption for surviving patients at 5-years. Additional exploratory undertaken by the ERG suggests that that the costs of standard care chemotherapy, the post-HSCT survival probabilities and the utility value for the post-relapse state have only a minor impact on the ICER for blinatumomab versus standard care. Conversely, the cure fraction and the choice of parametric OS distribution have a significant impact on the ICER for blinatumomab versus standard care. Within the ERG's exploratory analysis of alternative RFS and OS functions, the ICER for blinatumomab versus standard care ranges from a lowest ICER of £25,783 per QALY gained (unrestricted Weibull non-mixture cure model) to a highest ICER of £63,265 per QALY gained (unrestricted Weibull). Across the full range of models considered, only the Weibull non-mixture cure model (unrestricted) and the Weibull mixture cure model (unrestricted) produce results in which the full range of ICERs are below £30,000 per QALY gained (irrespective of RFS model assumed). The clinical advisors' three preferred OS models (Generalised gamma [unrestricted], RCS Weibull [unrestricted] and Weibull mixture cure [unrestricted]) result in ICERs in the range £25,810 per QALY gained to £34,904 per QALY gained. The ERG notes that all analyses should be considered highly uncertain due to the absence of RCT evidence for blinatumomab versus standard care and a lack of evidence relating to long-term RFS and OS outcomes for patients treated with blinatumomab (including the timing of cure). The ERG further notes that no comparative clinical or economic evidence is available for the comparison of blinatumomab versus standard care chemotherapy in patients Ph+ MRD+ BCP-ALL or in MRD+ BCP-ALL patients in CR2+, or for the comparison of blinatumomab versus monitoring in patients who are unable to undergo HSCT or to tolerate chemotherapy.

The ERG does not believe that blinatumomab meets NICE's criteria for life-extending treatments given at the end of life.

8. REFERENCES

1. Amgen Ltd. Blinatumomab for acute lymphoblastic leukaemia for people with minimal residual disease activity in remission [ID1036]. Company's evidence submission to the National Institute for Health and Care Excellence. In. Cambridge, UK: Amgen; 2017.
2. Alvarnas JC, Brown PA, Aoun P, Ballen KK, Bellam N, Blum W, *et al.* Acute lymphoblastic leukemia. *J Natl Compr Canc Netw* 2012;10:858-914.
3. National Comprehensive Cancer Network. NCCN Clinical Practice Guidelines in Oncology: Acute Lymphoblastic Leukemia. Version 1.2014. 2014. https://www.nccn.org/professionals/physician_gls/f_guidelines.asp (Accessed 23 November 2017).
4. Goekbuget N, Kneba M, Raff T, Trautmann H, Bartram CR, Arnold R, *et al.* Adult patients with acute lymphoblastic leukemia and molecular failure display a poor prognosis and are candidates for stem cell transplantation and targeted therapies. *Blood* 2012;120:1868-76.
5. Thomas X, Boiron J-M, Huguet F, Dombret H, Bradstock K, Vey N, *et al.* Outcome of treatment in adults with acute lymphoblastic leukemia: analysis of the LALA-94 trial. *Journal of clinical oncology* 2004;22:4075-86.
6. Moorman AV, Harrison CJ, Buck GA, Richards SM, Secker-Walker LM, Martineau M, *et al.* Karyotype is an independent prognostic factor in adult acute lymphoblastic leukemia (ALL): analysis of cytogenetic data from patients treated on the Medical Research Council (MRC) UKALLXII/Eastern Cooperative Oncology Group (ECOG) 2993 trial. *Blood* 2007;109:3189-97.
7. Oriol A, Vives S, Hernández-Rivas J-M, Tormo M, Heras I, Rivas C, *et al.* Outcome after relapse of acute lymphoblastic leukemia in adult patients included in four consecutive risk-adapted trials by the PETHEMA Study Group. *Haematologica* 2010;95:589-96.
8. Chiaretti S, Vitale A, Cazzaniga G, Orlando SM, Silvestri D, Fazi P, *et al.* Clinico-biologic features of 5202 acute lymphoblastic leukemia patients enrolled in the Italian AIEOP and GIMEMA Protocols and stratified in age-cohorts. *Haematologica* 2013:haematol. 2012.080432.
9. Fielding AK, Richards SM, Chopra R, Lazarus HM, Litzow MR, Buck G, *et al.* Outcome of 609 adults after relapse of acute lymphoblastic leukemia (ALL); an MRC UKALL12/ECOG 2993 study. *Blood* 2007;109:944-50.
10. Moorman AV, Chilton L, Wilkinson J, Ensor HM, Bown N, Proctor SJ. A population-based cytogenetic study of adults with acute lymphoblastic leukemia. *Blood* 2010;115:206-14.
11. Radich JP. Philadelphia chromosome-positive acute lymphocytic leukemia. *Hematol Oncol Clin North Am* 2001;15:21-36.
12. Fielding AK. Current treatment of Philadelphia chromosome-positive acute lymphoblastic leukemia. *Hematology Am Soc Hematol Educ Program* 2011:231-7.
13. Hoelzer D, Bassan R, Dombret H, Fielding A, Ribera J, Buske C. Acute lymphoblastic leukaemia in adult patients: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Annals of Oncology* 2016;27:v69-v82.
14. Campana D. Minimal residual disease in acute lymphoblastic leukemia. *ASH Education Program Book* 2010;2010:7-12.
15. Bassan R, Spinelli O, Oldani E, Intermesoli T, Tosi M, Peruta B, *et al.* Improved risk classification for risk-specific therapy based on the molecular study of minimal residual disease (MRD) in adult acute lymphoblastic leukemia (ALL). *Blood* 2009;113:4153-62.

16. Holowiecki J, Krawczyk-Kulis M, Giebel S, Jagoda K, Stella-Holowiecka B, Piatkowska-Jakubas B*, et al.* Status of minimal residual disease after induction predicts outcome in both standard and high-risk Ph-negative adult acute lymphoblastic leukaemia. The Polish Adult Leukemia Group ALL 4-2002 MRD Study. *British Journal of Haematology* 2008;142:227-37.

17. Amgen Ltd. Blinatumomab for acute lymphoblastic leukaemia for people with minimal residual disease activity in remission. Response to clarification questions. In. Cambridge, UK: Amgen; 2017.

18. Berry DA, Zhou S, Higley H, Mukundan L, Fu S, Reaman GH*, et al.* Association of Minimal Residual Disease With Clinical Outcome in Pediatric and Adult Acute Lymphoblastic Leukemia: A Meta-analysis. *JAMA oncology* 2017;3:e170580.

19. UKALL14. A randomized trial for adults with newly diagnosed acute lymphoblastic leukaemia. 2012. 201210/01/2018).

20. Bruggemann M, Raff T, Flohr T, Goekbuget N, Nakao M, Droese J*, et al.* Clinical significance of minimal residual disease quantification in adult patients with standard-risk acute lymphoblastic leukemia. *Blood* 2006;107:1116-23.

21. Amgen Inc. Data on File: Minimal residual disease testing patterns and healthcare utilization for adults with B-precursor acute lymphoblastic leukemia - results of a physician survey in the UK. In; 2017.

22. National Institute for Health and Care Excellence. Blinatumomab for acute lymphoblastic leukaemia for people with minimal residual disease activity in remission: final scope. In. London, UK.: NICE; 2017.

23. van Dongen JJ, van der Velden VH, Brüggemann M, Orfao A. Minimal residual disease diagnostics in acute lymphoblastic leukemia: need for sensitive, fast, and standardized technologies. *Blood* 2015;125:3996-4009.

24. European Medicines Agency. Summary of Product Characteristics - blinatumomab. EMA. https://www.medicines.org.uk/emc/product/5064 (Accessed 12/01/2018).

25. Royal Pharmaceutical Society. Medicines Complete (BNF). 2017.

26. National institute for Health and Care Excellence. Guide to the methods of technology appraisal. London, UK: NICE; 2013.

27. Higgins JPT, Green S. Cochrane Handbook for Systematic Reviews of Interventions. Version 5.1.0 [updated March 2011] edn: The Cochrane Collaboration; 2011.

28. Sterne JAC, Hernán MA, Reeves BC, Savović J, Berkman ND, Viswanathan M*, et al.* ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *BMJ* 2016;355.

29. Dalziel K, Round A, Stein K, Garside R, Castelnuovo E, Payne L. Do the findings of case series studies vary significantly according to methodological characteristics? *Health technology assessment (Winchester, England)* 2005;9:iii-iv, 1-146.

30. Centre for Reviews and Dissemination. Systematic Reviews: CRD's Guidance for Undertaking Reviews in Healthcare. 3rd Edition edn. York: CRD, University of York; 2009.

31. Amgen Inc. Clinical Study Report: MT103-203 Addendum 1 (Key Secondary Analysis). In; 2016.

32. Amgen Inc. Clinical Study Report: MT103-202. In. Cambridge, UK; 2013.

33. Amgen Inc. Observational Research Study Report: A Retrospective Analysis of Hematological Relapse Free Survival (RFS) and Overall Survival (OS) in Adult Patients with Philadelphia-Negative B-Precursor Acute Lymphoblastic Leukemia (ALL) in Complete Hematological Remission with Minimal Residual Disease (MRD). . In; 2017.

34. Evans SR. Clinical trial structures. *Journal of experimental stroke & translational medicine* 2010;3:8-18.

35. Chambers D, Rodgers M, Woolacott N. Not only randomized controlled trials, but also case series should be considered in systematic reviews of rapidly developing technologies. *Journal of clinical epidemiology* 2009;62:1253-60.e4.

36. Jüni P, Altman DG, Egger M. Assessing the quality of controlled clinical trials. *BMJ* 2001;323:42-6.

37. Sedgwick P. Retrospective cohort studies: advantages and disadvantages. *BMJ : British Medical Journal* 2014;348.

38. Amgen Inc. Study Protocol: A confirmatory multicenter, single-arm study to assess the efficacy, safety, and tolerability of the BiTE antibody blinatumomab in adult patients with minimal residual disease (MRD) of B-precursor acute lymphoblastic leukemia. In; 2010.

39. Goekbuget N, Dombret H, Bonifacio M, Reichle A, Graux C, Havelange V, *et al.* BLAST: A Confirmatory, Single-Arm, Phase 2 Study of Blinatumomab, a Bispecific T-Cell Engager (BiTE®) Antibody Construct, in Patients with Minimal Residual Disease B-Precursor Acute Lymphoblastic Leukemia (ALL). *Blood* 2014;124:379.

40. Gökbuget N, Zugmaier G, Klinger M, Kufer P, Stelljes M, Viardot A, *et al.* Long-term relapse-free survival in a phase 2 study of blinatumomab for the treatment of patients with minimal residual disease in B-lineage acute lymphoblastic leukemia. *Haematologica* 2017;102:e132-e5.

41. Pulte D, Jansen L, Castro FA, Brenner H. Changes in the survival of older patients with hematologic malignancies in the early 21st century. *Cancer* 2016;122:2031-40.

42. Fayers PM AN, Bjordal K, et al. . . The EORTC QLQ-C30 Scoring Manual (3rd Edition). 2001.

43. European Medicines Agency. EPAR for Blincyto (blinatumomab). http://www.ema.europa.eu/ema/index.jsp?curl=pages/medicines/human/medicines/003731/human_med_001921.jsp&mid=WC0b01ac058001d124 (Accessed 28th January).

44. Amgen Inc. 203-148-MRD-propensity-score-analysis-report-att-updated-final. In; 2017.

45. Faria R, Hernandez Alava M, Manca A, Wailoo A. NICE DSU technical support document 17: the use of observational data to inform estimates of treatment effectiveness for technology appraisal: methods for comparative individual patient data. *Sheffield: NICE Decision Support Unit* 2015.

46. Cole SR, Hernan MA. Adjusted survival curves with inverse probability weights. *Computer methods and programs in biomedicine* 2004;75:45-9.

47. National institute for Health and Care Excellence. Inotuzumab ozogamicin for treating relapsed or refractory B-cell acute lymphoblastic leukaemia. London; 2017.

48. UK Stem Cell Strategy Oversight Committee. NHS Blood and Transplant. Unrelated Donor Stem Cell Transplantation in the UK. Effective Affordable Sustainable; 2014.

49. van Agthoven M, Groot MT, Verdonck LF, Lowenberg B, Schattenberg AV, Oudshoorn M, *et al.* Cost analysis of HLA-identical sibling and voluntary unrelated allogeneic bone marrow and peripheral blood stem cell transplantation in adults with acute myelocytic leukaemia or acute lymphoblastic leukaemia. *Bone marrow transplantation* 2002;30:243-51.

50. Kind P, Hardman G, Macran S. UK Population Norms for EQ-5D. University of York; 1999.

51. Kurosawa S, Yamaguchi T, Mori T, Kanamori H, Onishi Y, Emi N, *et al.* Patient-reported quality of life after allogeneic hematopoietic cell transplantation or chemotherapy for acute leukemia. *Bone marrow transplantation* 2015;50:1241-9.

52. Commercial Medicines Unit. Drugs and Pharamaceutical electronic market information (eMit). 2017. https://www.gov.uk/government/publications/drugs-and-pharmaceutical-electronic-market-information-emit (Accessed 11/12/2017).

53. National Institute for Health and Care Excellence. Blinatumomab for treating Philadelphiachromosome-negative relapsed or refractory acute lymphoblastic leukaemia. London, UK: NICE; 2017.

54. Addicott R, Dewar S. Improving choice at end of life: a descriptive analysis of the impact and costs of the Marie Cure Delivering Choice Programme in Lincolnshire. . London, UK: King's Fund; 2008.

55. Marie Curie Cancer Care. Understanding the cost of end of life care in different settings. 2012. https://www.mariecurie.org.uk/globalassets/media/documents/commissioning-our-services/publications/understanding-cost-end-life-care-different-settingspdf (Accessed 11/01/2018).

56. Department of Health. NHS Reference Costs 2015 to 2016. 2016. https://www.gov.uk/government/publications/nhs-reference-costs-2015-to-2016 (Accessed 10/01/2018).

57. Collett D. Modelling survival data in medical research: CRC press; 2015.

58. Ara R, Brazier JE. Populating an economic model with health state utility values: moving toward better practice. *Value in Health* 2010;13:509-18.

59. Drummond M, Sculpher M, Torrance G, O'Brien B, Stoddart G. Methods for the Economic Evaluation of Health Care Programmes. 3rd edn. New York: Oxford University Press; 2005.

60. Weinstein MC, O'Brien B, Hornberger J, Jackson J, Johannesson M, McCabe C, *et al.* Principles of good practice for decision analytic modeling in health-care evaluation: report of the ISPOR Task Force on Good Research Practices--Modeling Studies. *Value in Health* 2003;6:9-17.

61. Amgen Inc. ALL epidemiology estimates. *Company evidence submission template for blinatumomab for acute lymphoblastic leukaemia for people with minimal residual disease activity in remission [ID1036]*, 2017.

62. Goldstone AH, Richards SM, Lazarus HM, Tallman MS, Buck G, Fielding AK, *et al.* In adults with standard-risk acute lymphoblastic leukemia, the greatest benefit is achieved from a matched sibling allogeneic transplantation in first complete remission, and an autologous transplantation is less effective than conventional consolidation/maintenance chemotherapy in all patients: final results of the International ALL Trial (MRC UKALL XII/ECOG E2993). *Blood* 2008;111:1827-33.

9. APPENDICES

Appendix 1: Technical appendix detailing methods for applying the ERG's exploratory analyses within the company's model

Exploratory analysis 1 – correction of model errors

ERG exploratory analysis 1 corrects five errors which were not previously addressed within the company's updated model.

1. Annual general population mortality rate applied for 1-year intervals defined according to time since model entry, rather than according to patient age. Not corrected in company's updated model.
2. Risk of all-cause death exceeds 1.0 for males patients aged 95 years and older and female patients aged 97 years and older. Corrected in company's updated model.
3. Minor programming errors in formula used to calculate receipt of HSCT at 2 years. Not corrected in company's updated model.
4. Incorrect formula offset in discounting cost of other inpatient visits post-relapse in the blinatumomab group. Not corrected in company's updated model.
5. Post-relapse HSCTs assumed to occur after the 5-year time point; inconsistent with the rest of the model structure. Corrected in company's updated model.
6. Inappropriate application of utility decrement due to proximity to death. Not corrected in company's updated model.
7. Incorrectly discounting of HSCT costs due to the use of approximate =LOOKUP() functions used to calculate the discount rate for receipt of HSCT. Not corrected in company's updated model.

1. Correct mortality lookup error

a. Open the model ID1026 Blin MRD Ph- B ALL_Updated CEM.xlsb
b. Go to "Blin Calc" worksheet
c. Go to cell AW9
d. Type the formula "=VLOOKUP(ROUNDDOWN(AT9,0),$AL$9:$AR$91,4,TRUE)"
e. Copy the formula down column AW
f. Go to cell AX9
g. Type the formula "=MAX(VLOOKUP(ROUNDDOWN(AT9,0)+1,$AL$9:$AR$91,4,TRUE),0)"
h. Copy the formula down column AX
i. Go to "SOC Calc" worksheet
j. Go to cell AW9
k. Type the formula "=VLOOKUP(ROUNDDOWN(AT9,0),'SOC Calc'!$AL$9:$AP$91,4,TRUE)"
l. Copy the formula down column AW
m. Go to cell AX9

n. Type the formula “=MAX(VLOOKUP(ROUNDDOWN(AT9,0)+1,'SOC Calc'!$AL$9:$AP$91,4,TRUE),0)”
o. Copy the formula down column AX9

3. HSCT programming error

a. Go to “Blin Calc” worksheet
b. Go to cell GZ13
c. Type the formula “=GY12*(hsct.pctmo24/hsct.pctmo13)”
d. Copy the formula down column GZ
e. Go to cell HQ13
f. Type the formula “=HP12*(hsct.pctmo24/hsct.pctmo13)”
g. Copy the formula down column HQ
h. Go to “SOC Calc” worksheet
i. Go to cell GZ13
j. Type the formula” =GY12*(hsct.pctmo24/hsct.pctmo13)”
k. Copy the formula down column GZ
l. Go to cell HQ 13
m. Type the formula “=HP12*(hsct.pctmo24/hsct.pctmo13)”
n. Copy the formula down column HP

4. Correct cost discounting formula for blinatumomab

a. Go to “Blin Calc” worksheet
b. Go to cell E27
c. Type the formula “=SUMPRODUCT($GC$9:$GC$3138,$M$9:$M$3138,R9:R3138)”

6. Correction proximity to death decrement

a. Go to “Blin Calc” worksheet
b. Go to cell JH9
c. Type the formula “=IF(IO9<model.term_util_end,MIN(J9,0.5),0)”
d. Copy the formula down column JH
e. Go to cell JI9
f. Type the formula “=JG9*util.term*JH9”
g. Copy the formula down column JI
h. Go to “SOC Calc” worksheet
i. Go to cell JH9
j. Type the formula “=IF(IO9<model.term_util_end,MIN(J9,0.5),0)”
k. Copy the formula down column JH
l. Go to cell JI9
m. Type the formula “=JG9*util.term*JH9”
n. Copy the formula down column JI

7. Incorrectly discounting of HSCT costs

a. Go to “Blin Calc” worksheet
b. Go to cell GS9
c. Type the formula “=1/(1+_input_model.discount_cost)^ROUNDDOWN(GQ9,0)”
d. Copy the formula down column GS
e. Go to the “SOC Calc” worksheet

f. Go to cell GS9
g. Type the formula “=1/(1+_input_model.discount_cost)^ROUNDDOWN(GQ9,0)”
h. Copy the formula down column GS

Exploratory analysis 2 – application of cure point at 5-years

a. Go to worksheet “Settings” cell G26.
b. Select “switch” from the drop-down menu.

Exploratory analysis 3 – ERG-preferred analysis

Combine exploratory analysis 1 and 2.

All subsequent exploratory analyses are based on this version of the model.

Exploratory analysis 4 - impact of alternative standard care chemotherapy costs

a. Go to the “Cost Inputs” sheet
b. Go to cell F84
c. Type the formula “=29.26*2”
d. Go to cell G84
e. Type the formula “=0.41*2”
f. Go to cell H84
g. Type the formula “=49.15*2”
h. Go to cell I84
i. Type the formula “=4.39*2”
j. Go to cell J84
k. Type the formula “=6.63*2”

Exploratory analysis 5 - alternative post-HSCT survival probabilities

a. Go to the “Blin calc” worksheet
b. Insert a new column GR
c. Go to cell GR8 type Age(years)
d. Go to cell GR9 use the formula “=ROUNDDOWN($K$9+GQ9,0)”
e. Copy this formula down column GR
f. Insert three new columns GS, GT, and GU
g. Label column GS “ probability of death (1st 6 months)”
h. Label Column GT probability of death (future months)
i. Label column GU rate of death
j. Go to cell AS8 type “Gender weighted probability of dying between ages”
k. Go to cell AS9 and type the formula “=model.pct_male*AM9+(1-model.pct_male)*AN9”
l. Copy this formula down column AS
m. Go to cell GU and type the following formula “=-(LN(1-VLOOKUP(GR9,$AL$9:$AS$91,8,FALSE))/365.25)”
n. Copy this formula down GU
o. Go to cell GS9 and type the formula =IFERROR(0.07+1-EXP(-GU9*((365.25/2)-100)),100%)”
p. Copy this formula down column GS
q. Go to cell GT9 and type the formula “=IFERROR(1-EXP(-GU9*((365.25/2))),100%)”
r. Copy this formula down column GT
s. Go to Cell GZ9 and type the formula ”=GY9*(1-$GS9)”
t. Copy this formula down column GZ

u. Go to cell HA10 type the formula "=GZ9*(1-$GT10)"
v. Copy down
w. Copy cell HA10
x. Paste the formula into cells HB11, HC12, HD13, HE14, HF15, HG16, HH17, HI18
y. Copy the formulae down columns HA, HB, HC, HD, HE, HF, HG, HH
z. Go to cell HJ19 and type the formula " =(HI18+HJ18)*(1-$GT19)"
aa. Copy down column HJ
bb. Select cells GZ9:HJ129
cc. Copy the cells
dd. Select cell HQ9
ee. Paste the formulae
ff. Go to the SOC Calc worksheet
gg. Go to cell GV9 and type the formula "=GU9*(1-'Blin Calc'!$GS9)"
hh. Copy this formula down column GV.
ii. Go to cell GW10 and type the formula "=GV9*(1-'Blin Calc'!$GT10)"
jj. Copy this formula down column GW
kk. Copy cell GW10
ll. Paste the formula into cells GX11, GY12, GZ13, HA14, HB15, HC16, HD17, HE18
mm. Copy down columns GX, GY, GZ, HA, HB, HC, HD, HE
nn. Go to cell HF19 and type "=(HE18+HF18)*(1-'Blin Calc'!$GT19)"
oo. Copy this formula down column HF
pp. Copy cells GV9:HF129
qq. Select cell HM9
rr. Paste the formulae

Exploratory analysis 6 - alternative cure fractions for the standard care group

a. Go to worksheet "SOC Calc" cell CM15
b. Apply alternative cure fractions

Exploratory analysis 7 - alternative post-relapse utilities

a. Go to worksheet "Utility Inputs" cell F18
b. Apply alternative post-relapse utility values

Exploratory analysis 8 - Exploration of the impact of alternative parametric RFS and OS models

Run macro as per instructions provided by the company using ERG-preferred model

Appendix 2: Blinatumomab for acute lymphoblastic leukaemia for people with minimal residual disease activity in remission - Model selection exercise

Background information

Within their health economic model, the company has fitted a range of parametric survivor functions to time-to-event outcomes (overall survival [OS] and relapse-free survival [RFS]) for patients with Ph- disease with CR1 from the BLAST study and the ATT-weighted historical control study in order to extrapolate beyond the duration of the empirical studies. These survival curves influence both the costs and the health gains predicted by the company's model. We have some concerns regarding how the company has selected their preferred survival curves for use in the model, particularly with respect to the plausibility of the extrapolated portion of the curve. Our main concern is surrounding OS, as this is a key driver of the cost-effectiveness of blinatumomab.

Based on clinical advice, we believe that it would be broadly appropriate to assume that patients who have not relapsed within 5-years are cured. For simplicity, we have assumed the same to be true with respect to OS, although we note that some relapsed patients may achieve cure as a consequence of downstream treatments (e.g. HSCT received post-relapse), hence the time at which cure manifests may be slightly later for OS than RFS.

We have plotted the Kaplan-Meier curves from the BLAST and historical control studies and have overlaid these with a range of long-term potential OS survivor functions (see Figures 1 and 2). As a consequence of the assumption of cure at 5-years, all models are based on the company's statistical model projections for up to 5-years; the survivor function is then applied using uplifted general population mortality rates thereafter. The model assumes a population starting age of roughly 45 years.

Your task

We now need to choose which curve is likely to be most appropriate for OS. We would like you to look at the fitted curves presented in Figure 1 (blinatumomab OS) and Figure 2 (standard care OS) and to fill in the responses to questions on pages 4 and 5 to indicate which of the curves you consider to be the most clinically plausible and to state your reasons why. In completing this exercise, please consider both how well the curve appears to fit the observed data as well as the clinical plausibility of the extrapolation beyond the observed period. To do this you may wish to think about:

- The distance between the smooth parametric curves and the stepped Kaplan-Meier function (note that the end of the Kaplan-Meier curve is very uncertain)
- The proportion of patients you would expect to achieve a cure by 5-years
- The probability of surviving at different timepoints in each treatment group

We note that several of the curves appear to be very similar. If you wish to select multiple preferred curves, please do so.

Figure 1: Comparison of alternative OS survivor functions (including 5-year cure assumption) - blinatumomab

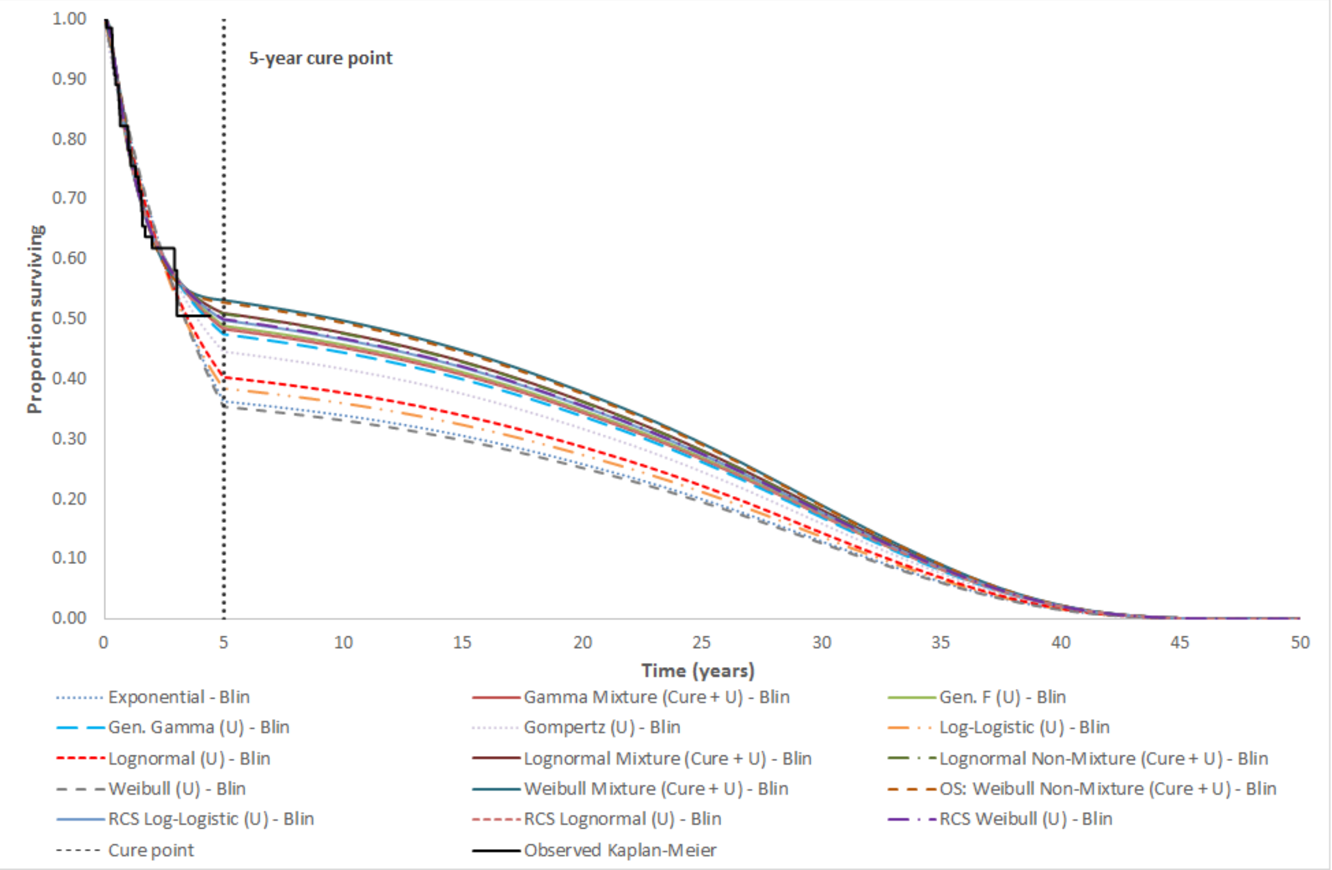
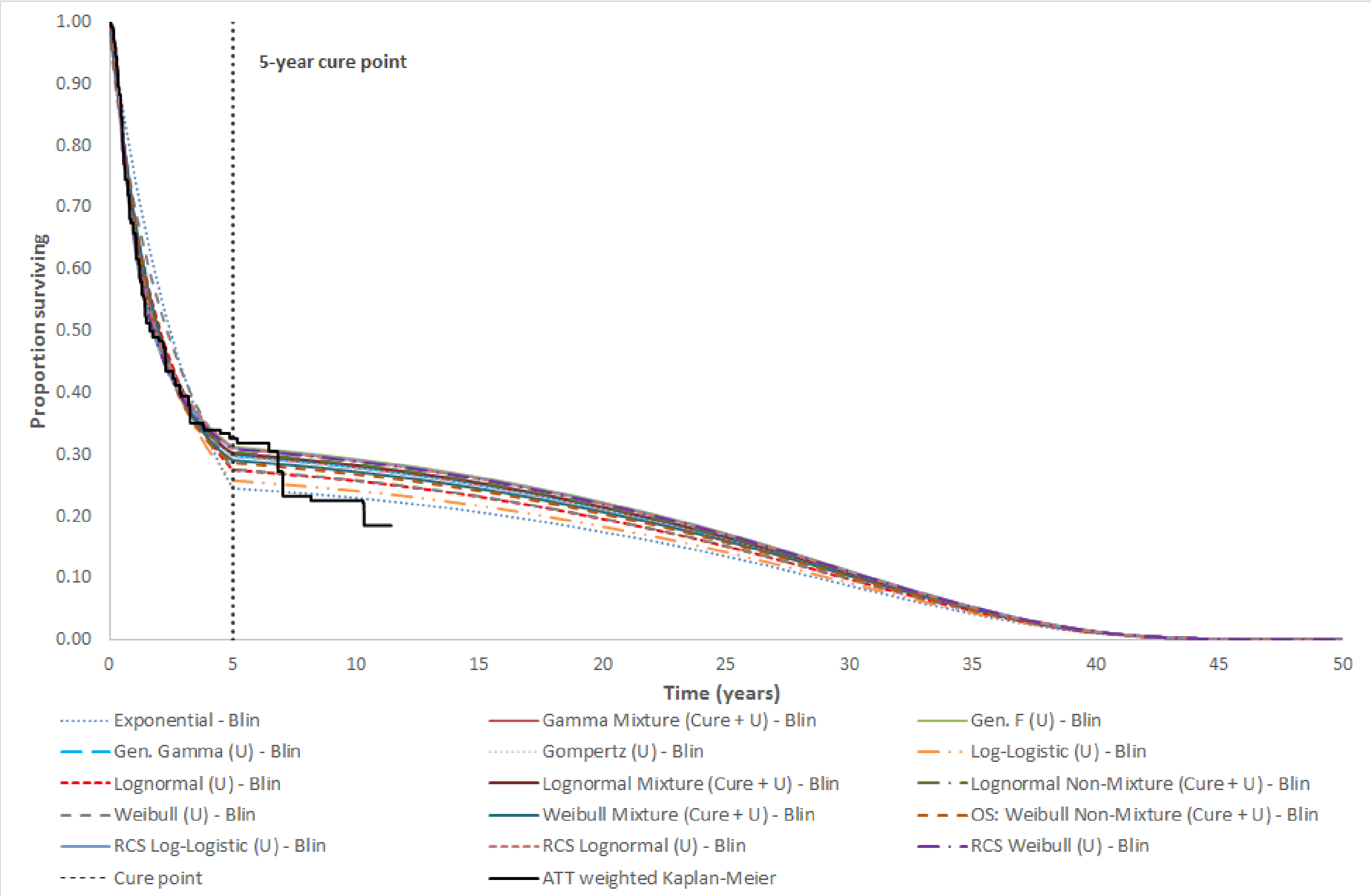


Figure 2: Comparison of alternative OS survivor functions (including 5-year cure assumption) – standard care



Clinicians' responses

QUESTION 1. Do you think it is reasonable to apply the cure point at 5-years? Or should we assume a later timepoint for OS?

RESPONSE 1:

Blinatumomab group (Please refer to Figure 1)

QUESTION 2. Which is your preferred OS function for the blinatumomab group?

RESPONSE 2:

QUESTION 3. Please state why this is your preferred function

RESPONSE 3:

QUESTION 4. Which other functions would you consider to be plausible?

RESPONSE 4:

Standard care group (Please refer to Figure 2)

QUESTION 5. Which is your preferred OS function for the standard care group?

RESPONSE 5:

QUESTION 6. Please state why this is your preferred function

RESPONSE 6:

QUESTION 7. Which other functions would you consider to be plausible?

RESPONSE 7:
