NICE National Institute for Health and Care Excellence

Single Technology Appraisal

Treosulfan with fludarabine for malignant disease before allogeneic stem cell transplant [ID1508]

Committee Papers

NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE

SINGLE TECHNOLOGY APPRAISAL

Treosulfan with fludarabine for malignant disease before allogeneic stem cell transplant [ID1508]

Contents:

The following documents are made available to consultees and commentators:

The **final scope and final stakeholder list** are available on the NICE website.

1. **Company submission** from Medac
2. **Clarification questions and company responses**
3. **Expert personal perspectives** from:
   a. Dr Amit Patel, Consultant/Programme Director, Stem Cell Transplantation and Cellular Therapy - clinical expert, nominated by Medac
   b. Caitlin Farrow, Head of Policy & Public Affairs - patient expert, nominated by Anthony Nolan
   c. Alan Tindale – patient expert, nominated by Anthony Nolan
4. **Evidence Review Group report** prepared by Kleijnen Systematic Reviews
5. **Evidence Review Group report – factual accuracy check**
6. **Technical engagement response** from Medac
7. **Final Technical Report**

Any information supplied to NICE which has been marked as confidential, has been redacted. All personal information has also been redacted.

NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE

Single technology appraisal

Treosulfan with fludarabine for malignant disease before allogeneic stem cell transplant [ID1508]

Document B

Company evidence submission

May 2019

File name	Version	Contains confidential information	Date
ID1508 Document B – Company evidence submission – medac treosulfan	Final2	Yes	31 May 2019

Introduction to this document

This document represents the medac Pharma's evidence submission for the review of ID1508: Treosulfan with fludarabine for malignant disease before allogeneic stem cell transplant.

Table of contents

List of tables

List of figures

Abbreviations

Abbreviation	Definition
AD	Cytarabine and daunorubicin
AE	Adverse event
aGvHD	Acute graft versus host disease
AIC	Akaike information criterion
ALL	Acute lymphoblastic leukaemia
AlloHSCT	Allogeneic haematopoietic stem cell transplantation
AML	Acute myeloid leukaemia
ASBMT	American Society for Blood and Marrow Transplantation
ATG	Anti-thymocyte globulin
AutoHSCT	Autologous hematopoietic stem cell transplantation
BIC	Bayesian information criteria
BM	Bone marrow
BMF	Bone marrow failure
BMT	Bone marrow transplantation
BMTCTN	Blood and Marrow Transplant Clinical Trial Network;
BNF	British National Formulary
BSA	Body surface area
BSBMT	British Society of Blood and Marrow Transplantation
BSC	Best supportive care
BU	Busulfan
BU/CY	Busulfan/cyclophosphamide
BU/FLU	Busulfan/fludarabine
BW	Body weight
CEAC	Cost-effectiveness acceptability curve
cGvHD	Chronic graft versus host disease
CI	Confidence interval
CIBMTR	Center for International Blood and Marrow Transplant Research
CML	Chronic myeloid leukaemia
CMML	Chronic myelomonocytic leukaemia
CNS	Central nervous system
ConCR	Conventional care regimens
COPD	Chronic obstructive pulmonary disease
CR	Complete remission
CR1, CR2	First complete remission, second complete remission
CRF	Case report form
CRO	Contract research organisation
CRP	C-reactive protein
CSR	Clinical study report
CT	Chemotherapy
CTC	Common Terminology Criteria
CTCAE	Common Terminology Criteria for Adverse Events
CTR	Clinical trial report
CY	Cyclophosphamide
DCE	Discrete choice experiment
DFS	Disease-free survival
DLI	Donor lymphocyte infusion
DMC	Data monitoring committee

DNA	Deoxyribonucleic acid
DSU	Decision support unit
EBMT	European Society for Blood and Marrow Transplantation
EBV	Epstein-Barr virus
ECG	Electrocardiogram
EFS	Event-free survival
ELN	European Leukaemia Network
EMEA	European Medicines Agency
EORTC QLQ-C30	European Organization for the Research and Treatment of Cancer Quality of Life Questionnaire-C30
EPAR	European Public Assessment Report
EQ-5D	EuroQoL 5-dimension questionnaire
EQ-5D-3L	EuroQoL 5 dimension questionnaire, 3 level
ERG	Evidence Review Group
ETO	Etoposide
FACT-BMT	Functional Assessment of Bone Marrow Transplant
FACT-G	Functional Assessment of Cancer Therapy- General
FACT-Leu	Functional Assessment of Cancer Therapy- Leukaemia
FAS	Full analysis set
FBA	Fludarabine + oral busulfan + thymoglobulin
FEV_1	Forced expiratory volume in 1 second
FLU	Fludarabine
Flu/Mel	Fludarabine + melphalan
FTBI	Fludarabine + 2 Gy total body irradiation
GBP	Great British Pounds
G-CSF	Granulocyte colony-stimulating factor
GGT	Gamma glutamyl transferase
GM-CSF	Granulocyte-macrophage colony-stimulating factor
GRFS	GvHD-free survival
GvHD	Graft-versus-host disease
GvT	Graft-versus-tumour
Gy	Gray
HCHS	Hospital and Community Health Services
HCT	Haematopoietic cell transplantation
HCT-CI	Haematopoietic cell transplantation co-morbidity index
HERC	Health Economics Research Centre
HL	Hodgkin lymphoma
HLA	Human leukocyte antigen
HR	Hazard ratio
HRG	Healthcare Resource Group
HRQoL	Health-related quality-of-life
HSCT	Haematopoietic stem cell transplantation
HSOS	Hepatic sinusoidal obstruction syndrome
HSUV	Health state utilities value
i.v.	Intravenous
ICER	Incremental cost-effectiveness ratio
IMSF	International Myelodysplastic Syndromes Foundation
INMB	Incremental net monetary benefit
IPSS	International Prognostic Scoring System
IPSS-R	International Prognostic Scoring System-Revised

JMML	Juvenile myelomonocytic leukaemia
Kg	Kilogram
KOL	Key opinion leader
KPS	Karnofsky Performance Score (KPS)
LDC	Low dose chemotherapy
LVEF	Left ventricular ejection fraction
LYG	Life-year gain
MAC	Myeloablative conditioning
MAD	Median absolute error
MAE	Mean absolute error
MCM	Mixture-cure model
MDS	Myelodysplastic syndrome
MDS/MPN	Myelodysplastic/myeloproliferative neoplasms
MEL	Melphalan
mg	Milligram
MisMRD	Mismatched related donor
MM	Multiple myeloma
MRD	Matched related donors
MTD	Maximum tolerated dose
MTX	Methotrexate
MUD	Matched unrelated donor
N/A	Not applicable
N/n	Number
NCRI	National Cancer Research Institute
NHL	Non-Hodgkin Lymphoma
NHS	National Health Service
NICE	National Institute for Health and Care Excellence
NMA	Non-myeloablative
NMCM	Non-mixture cure models
NRM	Non-relapse mortality
OECD	Organisation for Economic Co-operation and Development
OLS	Ordinary least squares
OS	Overall survival
PB	Peripheral blood
PBSC	Peripheral blood stem cell
PCT	Procalcitonin
PFS	Progression-free survival
PID	Primary immunodeficiency
PO	Per os, orally
PPP	Power parity estimate
PPS	Per protocol set
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PSA	Probabilistic sensitivity analysis
PSS	Personal Social Services
PSSRU	Personal Social Services Research Unit
QALY	Quality-adjusted life-year
QLQ-C30	Quality of Life Questionnaire - Core Questionnaire
RBC	Red blood cell
RCT	Randomised clinical trial
RFS	Relapse-free survival

rHU-KGF	Recombinant human keratinocyte growth factor
RI	Relapse incidence
RIC	Reduced intensity conditioning
RMSE	Root mean squared error
RPE	Relative prediction error
RR	Relapse rate
RTC	Reduced toxicity conditioning
SAA	Severe aplastic anaemia
SAEs	Serious adverse events
SAP	Statistical Analysis Plan
SAR	Serious adverse reactions
SAS	Safety analysis set
SCID	Severe combined immune deficiency
SD	Standard deviation
SDC	Standard dose chemotherapy.
SE	Standard error
SF-12	Short form-12
SF-36	Short form-36
SLR	Systematic literature review
SmPC	Summary of product characteristics
SMR	Standardised mortality ratio
STA	Single technology appraisal
TA	Technology appraisal
TBI	Total body irradiation
TBI/CY	Total body irradiation + cyclophosphamide
TEAE	Treatment emergent adverse event
TREO	Treosulfan
TREO/FLU	Treosulfan + fludarabine
TRM	Transplantation-related mortality
TT	Thiotepa
TTO	Time trade-off
UCB	Umbilical cord blood
ULN	Upper limit of normal
VAS	Visual analogue scale
VOD	Veno-occlusive disease
vs	Versus
WHO	World Health Organization
WTP	Willingness to pay

Glossary

Haematopoietic stem cells: Rare subpopulation of haematopoietic cells, mostly residing in the bone marrow, that reign at the top of the haematopoietic hierarchy. These cells are characterised by their ability to self-renew and to differentiate to form all blood cell types.

Chimerism: State in bone marrow transplantation in which bone marrow and host cells exist compatibly without signs of graft-versus-host disease.

Graft-versus-host disease (GvHD): A disease in which the immune cells from the donor attack the tissues and organs of the recipient and occurs in 40-60% of allogeneic transplantation patients, resulting in considerable morbidity and mortality.[1]

Non-relapse mortality (NRM): Non-relapse mortality is defined as the probability of dying in the absence of persisting disease or previous occurrence of relapse or graft failures.

Transplantation-related mortality (TRM): In accordance with European Society for Blood and Marrow Transplantation (EBMT) definitions, all deaths occurring due to one of the following main causes are considered as transplantation-related:

- GvHD
- Graft failure
- Cardiac toxicity
- Infection
- Epstein-Barr virus (EBV) proliferative disease
- Pulmonary toxicity
- Hepatic sinusoidal obstruction syndrome (HSOS)
- Other haematopoietic stem cell transplantation (HSCT)-related cause

B.1 Decision problem, description of the technology and clinical care pathway

B.1.1 Decision problem

This technology appraisal evaluates the clinical and cost-effectiveness of treosulfan as a conditioning treatment for malignant disease prior to allogeneic haematopoietic stem cell transplantation (alloHSCT) in paediatric and adult patients older than one month. This submission focuses on part of the technology's authorisation, as it excludes non-malignant disease in adults. The focus is narrower than the marketing authorisation because the National Institute for Health and Care Excellence (NICE) has indicated that the non-malignant disease element of the full marketing authorisation will be subject to a separate single technology appraisal (STA) submission, GID-TA10453.

The full decision problem is described in Table 1 along with any differences between this submission and the Final Scope published by NICE, and the rationale for these differences.

Table 1: The decision problem

	Final scope issued by NICE	**Decision problem addressed in the company submission**	**Rationale if different from the final NICE scope**
Population	Adults, children and young people with malignant disease that is in remission before allogenic haematopoietic stem cell transplantation	As final scope	
Intervention	Treosulfan with fludarabine	As final scope	
Comparator(s)	Conditioning treatments (either high dose or reduced intensity): • cyclophosphamide and total body irradiation • cyclophosphamide and busulfan • busulfan with fludarabine • established clinical management without	As final scope	

	Final scope issued by NICE	Decision problem addressed in the company submission	Rationale if different from the final NICE scope
	treosulfan with fludarabine.		
Outcomes	The outcome measures to be considered include: • overall survival • event-free survival • rates of relapse • success of stem cell transplantation (engraftment) • adverse effects of treatment • health-related quality of life.	The outcome measures to be considered include: • overall survival • event-free survival • rates of relapse • success of stem cell transplantation (engraftment) • adverse effects of treatment • health-related quality of life • non-relapse mortality (NRM).	NRM has been added as a significant reduction in NRM is the reason for the overall survival benefit and event free survival benefit observed with treosulfan-based conditioning.

B.1.2 Description of the technology being appraised

A draft version of the summary of product characteristics (SmPC) has been included in Appendix C. This document is subject to being updated until publication of the European Public Assessment Report which has not yet been published (May 2019).

Treosulfan (Trecondi ®, medac GmbH) is a water-soluble prodrug of a bifunctional alkylating agent. Due to its proven antileukaemic and immunosuppressive activity, treosulfan in combination with fludarabine and other agents has been developed as conditioning regimen prior to alloHSCT in adults and children with malignant disease.

Treosulfan is administered by intravenous infusion which should be supervised by a physician experienced in conditioning treatment and alloHSCT. Monitoring procedures outlined in the SmPC (see Appendix C) should be followed.[2]

Table 2: Technology being appraised

UK approved name and brand name	Treosulfan (Brand name Trecondi®)
Mechanism of action	Treosulfan is a prodrug of a bifunctional alkylating agent with cytotoxic activity to haematopoietic precursor cells. The activity of treosulfan is due to the

<table>
<tr><td></td><td>spontaneous conversion into a mono-epoxide intermediate and L-diepoxybutan. The epoxides formed alkylate nucleophilic centres of deoxyribonucleic acid (DNA) and are able to induce DNA cross-links which are considered responsible for the stem cell depleting and antineoplastic effects (SmPC).</td></tr>
<tr><td>Marketing authorisation/CE mark status</td><td>In December 2017, medac GmbH (medac) submitted a marketing authorisation application (EMEA/H/C/004751) for treosulfan in a centralised procedure, according to Regulation (EC) No. 726/2004. On 13 December 2018, the Committee for Medicinal Products for Human Use (CHMP) adopted a positive opinion, recommending the granting of a marketing authorisation for the medicinal product treosulfan, intended for the conditioning treatment prior to alloHSCT. Medac anticipate that the final European Commission decision will not be published until May 2019 with the European Public Assessment Report (EPAR) published shortly afterwards.
An appeal was submitted in March 2019 regarding the orphan status of treosulfan. [redacted]
An injected form of treosulfan has had marketing authorisation in the UK since January 1992 and is indicated for the palliative treatment of epithelial ovarian cancer.[3]</td></tr>
<tr><td>Indications and any restriction(s) as described in the summary of product characteristics (SmPC)</td><td>The anticipated full indication is “Treosulfan in combination with fludarabine is indicated as part of conditioning treatment prior to allogeneic haematopoietic stem cell transplantation (alloHSCT) in adult patients with malignant and non-malignant diseases, and in paediatric patients older than one month with malignant diseases.”</td></tr>
<tr><td>Method of administration and dosage</td><td>Administration of treosulfan should be supervised by a physician experienced in conditioning treatment followed by alloHSCT.[2]
Adults with malignant disease: Treosulfan is given in combination with fludarabine. The recommended dose and schedule of administration is as follows:
Treosulfan 10 g/m² body surface area (BSA) per day as two-hour i.v. infusion, given on three consecutive days (day -4, -3, -2) before stem cell infusion (day 0). The total treosulfan dose is 30 g/m²
Fludarabine 30 mg/m² BSA per day as a 0.5-hour i.v. infusion, given on five consecutive days (day -6, -5, -4, -3, -2) before stem cell infusion (day 0). The total fludarabine dose is 150 mg/m²
Treosulfan should be administered before fludarabine on days -4, -3, -2 (FT_{10} regimen).
Paediatric population: Treosulfan is given in combination with fludarabine, with thiotepa (intensified regimen; $FT_{10-14}TT$ regimen) or without thiotepa (FT_{10-14} regimen). The recommended dose and schedule of administration is as follows:
Treosulfan 10–14 g/m² BSA/day as two-hour i.v. infusion, given on three consecutive days (day -6, -5, -4) before stem cell infusion (day 0). The total treosulfan dose is 30–42 g/m². The dose of treosulfan should be adapted to the patient’s BSA (see table below).
Treosulfan dose based on patient BSA.<table><tr><th>Body surface area (m²)</th><th>Treosulfan dose (g/m²)</th></tr><tr><td>≤0.5</td><td>10.0</td></tr></table></td></tr>
</table>

	<table><tr><td>>0.5–1.0</td><td>12.0</td></tr><tr><td>>1.0</td><td>14.0</td></tr></table> • Fludarabine 30 mg/m² BSA/day as a 0.5-hour i.v. infusion, given on five consecutive days (day -7, -6, -5, -4, -3) before stem cell infusion (day 0). The total fludarabine dose is 150 mg/m². • Treosulfan should be administered before fludarabine. • Thiotepa (intensified regimen) 2 × 5 mg/kg, given as two i.v. infusions over 2-4 hours on day -2 before stem cell infusion (day 0).
Additional tests or investigations	No additional tests or investigations are needed for treatment eligibility outside of those required in clinical practice for patients with malignant disease requiring chemotherapy. After initiating treosulfan, patients would require no tests or investigations additional to those that would already be performed following treatment with standard intensive chemotherapy.
List price and average cost of a course of treatment	Treosulfan is priced at £53.83 (1g vial) and £208.03 (5g vial). Total treatment cost for treosulfan is £2,496.41 (including wastage). Fludarabine's list price is £147.07 (50mg powder for solution for injection vials) and the total treatment cost for fludarabine is £3,059 (including wastage).
Patient access scheme (if applicable)	Not applicable.

B.1.3 Health condition and position of the technology in the treatment pathway

Summary of the health condition

- Haematopoietic stem cell transplantation (HSCT) is a potentially curative therapy for more than 70 malignant and non-malignant diseases but the exact HSCT procedure requires a delicate balance between the patient characteristics, the risk of relapse/progression and the late effects of the procedure.
- In the UK, the most common malignant indications for allogeneic HSCT (alloHSCT) are acute myeloid leukaemia (AML; 36%), acute lymphoblastic leukaemia (ALL; 16%) and myelodysplastic syndrome (MDS) or myelodysplastic/myeloproliferative neoplasms (MDS/MPN; 13%).[4]
- HSCT involves several complex procedures and requires a high level of expertise, thus treatments are only administered within specialised centres, with around 55 of these centres in the UK performing 1,366 first alloHSCTs in 2017.[5,4]
- After selection of the patient and donor, patients undergoing an alloHSCT are prepared with chemotherapy alone or chemotherapy combined with radiotherapy, the so-called conditioning therapy.

Conditioning therapy is followed by transplantation of stem cells which if successful is termed engraftment. The patient then undergoes post-graft immunosuppression and post-transplant follow-up where complications such as infection and graft-versus-host disease (GvHD) are monitored and treated.

- While standard myeloablative conditioning (MAC) regimens generally lead to low relapse rates, they are associated with high treatment-related toxicity and transplant-related mortality (TRM).[6]
- Non-myeloablative (NMA) conditioning regimens which may result in only minimal cytopenias that do not require stem cell support, and reduced intensity conditioning (RIC) where the cytopenia is sufficient to require stem cell support, were developed for patients such as the elderly and those with comorbidities where myeloablative conditioning is not considered optimal and to minimise treatment-related toxicity, non-relapse mortality (NRM) and TRM
- Unfortunately, lower dose intensity conditioning usually comes with a higher risk of relapse.[6]

Summary of the treatment pathway and the position of treosulfan

- Standard conditioning regimens have not been established for the various indications; as a result, UK clinical practice for patients undergoing alloHSCT varies considerably depending on the underlying malignant disease, age and other co-morbidities.
- Treosulfan is a reduced-toxicity conditioning (RTC) regimen that aims to provide the efficacy of MAC regimens to patients who cannot tolerate such MAC regimens and would otherwise be treated with less efficacious RIC regimens i.e. patients aged ≥ 50 years at transplant and/or with a Haematopoietic Cell Transplantation-Comorbidity Index (HCT-CI) score >2 (indicating an intermediate to high risk of co-morbidities which influence NRM).[7]
- Treosulfan reduces treatment-related toxicity and relapse rates compared with conventional conditioning therapies.[8,9]

B.1.3.1 Disease overview and epidemiology

B.1.3.1.1 Introduction

HSCT is a potentially curative therapy for many life-threatening cancers and non-malignant disorders. Today, more than 70 malignant and non-malignant diseases are treated routinely with HSCT. The indications for HSCT vary according to disease categories and are influenced by factors such as cytogenetic abnormalities,

response to prior therapy, patient age and performance status, disease status, and - most importantly - availability of a suitable graft source. Finding the optimal indications for HSCT is a delicate balance between risk of relapse/progression and late effects.

Autologous HSCT (autoHSCT) uses the patient's own cells and is currently considered as a standard therapy for patients with multiple myeloma (MM), some forms of non-Hodgkin lymphomas (NHL), as well as relapsed Hodgkin lymphomas (HL). Allogeneic HSCT (alloHSCT) uses cells from a donor and is potentially curative for acute and chronic leukaemias, myelo-dysplastic syndromes (MDS), HL, NHL, and MM.[10]

AlloHSCT is the treatment of choice in adult and paediatric patients with high-risk AML in their first complete remission (CR1). In patients with standard or good risk features, alloHSCT is reserved for their second complete remission (CR2). AlloHSCT is the only curative option for patients with primary refractory or relapsed AML.[10–12]

Furthermore, alloHSCT is increasingly used in non-malignant disease. This includes primary immunodeficiencies, inborn errors of metabolism, haemoglobinopathies and bone marrow failure syndromes.[10] However, this is the focus of a separate single technology assessment and will not be part of this submission.

B.1.3.1.2 Common haematological indications for HSCT in adults and children in the UK

According to the latest BSBMT registry, the most common indications for an alloHSCT in the UK are acute myeloid leukaemia (AML; 36%), acute lymphoblastic leukaemia (ALL; 16%) and myelodysplastic syndrome (MDS) or myelodysplastic/myeloproliferative neoplasms (MDS/MPN; 13%).[4] These indications for alloHSCT in the UK are comparable to those in adults and children undergoing HSCT in Europe (AML 38%; ALL 16%; and MDS or MDS/MPN 15%).[13] The use of alloHSCT in patients with chronic myeloid leukaemia (CML) has declined rapidly since the introduction of highly potent tyrosine kinase inhibitors like imatinib.[14]

In children, the main indications for alloHSCT in Europe are ALL (ALL; 26%), primary immunodeficiency (PID; 16%), AML (14%), bone marrow failure (BMF; 12%), thalassaemia (9%), and MDS/MPS (8%).[15]

B.1.3.1.3 Patients requiring HSCT in the UK

The BSBMT registry report of 2017 gives the total number of HSCTs including first and subsequent transplants performed in the UK as 4,489.[4] This is comprised of 1,684 alloHSCTs (1,594 first alloHSCTs; 95%)) and 2,805 autoHSCTs (2,483 first autoHSCTs; 89%). Table 3 summarises the reasons for the first allo- and autoHSCTs.

Table 3: Summary of first allo- and autoHSCTs in the UK (2017)

Disease	AlloHSCTs	AutoHSCTs	Total HSCTs
Malignant disease	1,366 (86%)	2,415 (97%)	3,781 (93%)
Non-malignant disease such as anaemia, immune deficiencies etc.	228 (14%)	68 (3%)	296 (7%)
Total	1,594	2,483	4,077

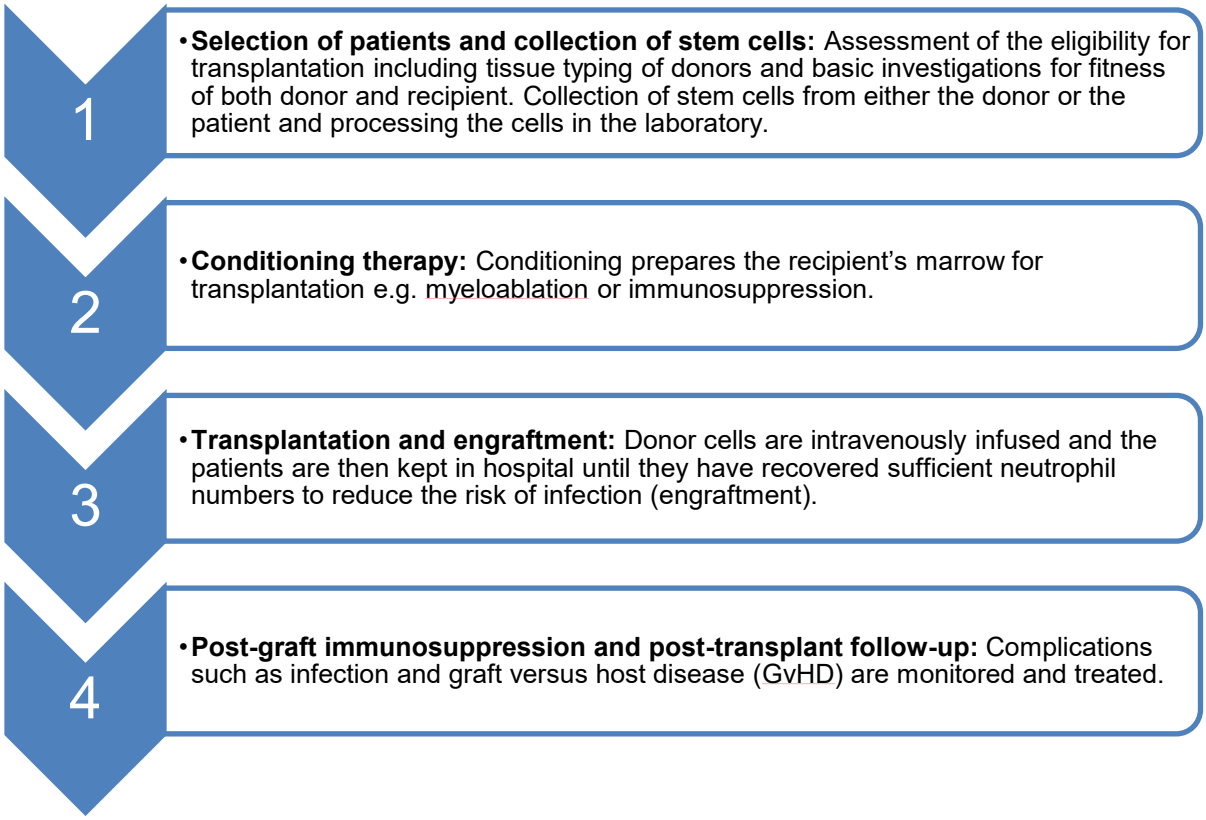
Source: BSBMT registry report (2017)[4]

B.1.3.1.4 Overview of HSCT procedure

HSCT involves several complex procedures and requires a high level of expertise, thus treatments are only administered within specialised centres, with just 55 of these centres listed by the British Society of Blood and Marrow Transplantation (BSBMT) in the UK in 2019.[5]

Typically, the procedure for HSCT can be divided into several consecutive steps as illustrated in Figure 1.[16,17]

Figure 1: Typical steps in HSCT procedure



B.1.3.2 Selection of patients and donors

Clinical decision-making for HSCT procedures is complex and includes several factors besides the underlying indication for transplantation. Decisions are greatly influenced by patient selection (e.g. interpretation of disease and patient-related factors influencing eligibility), HSCT approach (e.g. conditioning regimen, graft source and manipulation, donor selection) and HSCT potential complications (e.g. GvHD, organ toxicity, infections, relapse).[18]

Disease- and patient-related factors such as diagnosis, disease status (e.g. remission, refractory or relapsed disease), patient age and functional status and presence or absence of comorbid conditions are all key factors that influence the choice of alloHSCT approach and timing, as well as conditioning regimen.[19,20]

Finding the optimal indications for alloHSCT is a delicate balance between risk of relapse/progression and late effects. According to the European Society for Blood and Marrow Transplantation (EBMT), current strategies should focus on the concept

that patients with a high risk for TRM and a low disease risk should receive a different conditioning regimen from patients with a low risk for TRM and high risk disease.[21]

For patients with myelodysplastic syndrome (MDS) and chronic myelomonocytic leukaemia (CMML), an international expert panel active within the EBMT, the European Leukaemia Net (ELN), the Blood and Marrow Transplant Clinical Trial Network (BMTCTN) and the International Myelodysplastic Syndromes Foundation (IMSF) recommends that higher-risk patients with good performance status and no comorbidities are candidates for myeloablative regimens, whereas less fit patients or patients with comorbidities should be considered for RIC regimens.[20] A number of evidence reviews recommend a similar approach for patients with ALL, AML or MDS who are candidates for HSCT; key recommendations are summarised in Table 4.

Table 4: Summary of current recommendations on conditioning regimen intensity for HSCT

Recommendation	Disease	Evidence type	Reference
"Higher-risk patients with good performance status and no comorbidities are candidates for myeloablative regimens, whereas less fit patients or patients with comorbidities should be considered for RIC schedules (recommendation level C)"	MDS and CMML	International expert panel†	de Witte et al, 2017[20]
"Full-intensity conditioning HSCT should be considered for all fit high-risk ALL patients eligible for transplantation as it is supported by robust long-term follow-up data. The lack of long-term relapse-free survival (RFS) data precludes firm recommendations for RIC HSCT"	ALL	Evidence-based review	Dhawan and Marks, 2017[22]
"Myeloablative alloHSCT is an appropriate treatment for adult ALL patients in first complete remission for all disease risk groups. RIC regimens are appropriate only for ALL patients in remission who are unsuited for myeloablative conditioning"	ALL	Evidence-based review	Oliansky et al 2012[23]
"Patients at high risk for post-HSCT relapse with RIC and non-myeloablative regimens (e.g. advanced myeloid malignancies and aggressive lymphomas not in	Myeloid malignancies lymphomas	Evidence-based review	Gyurkocza et al, 2014[24]

Recommendation	Disease	Evidence type	Reference
remission) should be considered for more intense regimens”			
“RIC alloHSCT is a reasonable option for high-risk older patients and for younger AML patients with medical comorbidities who achieve a first or subsequent remission”	AML	Evidence-based review	Hamadani et al, 2011[8]
“Young and fit patients up to the age of 50–55 years should receive full-dose conditioning”	AML and MDS	Evidence-based review	Finke and Nagler, 2007[25]

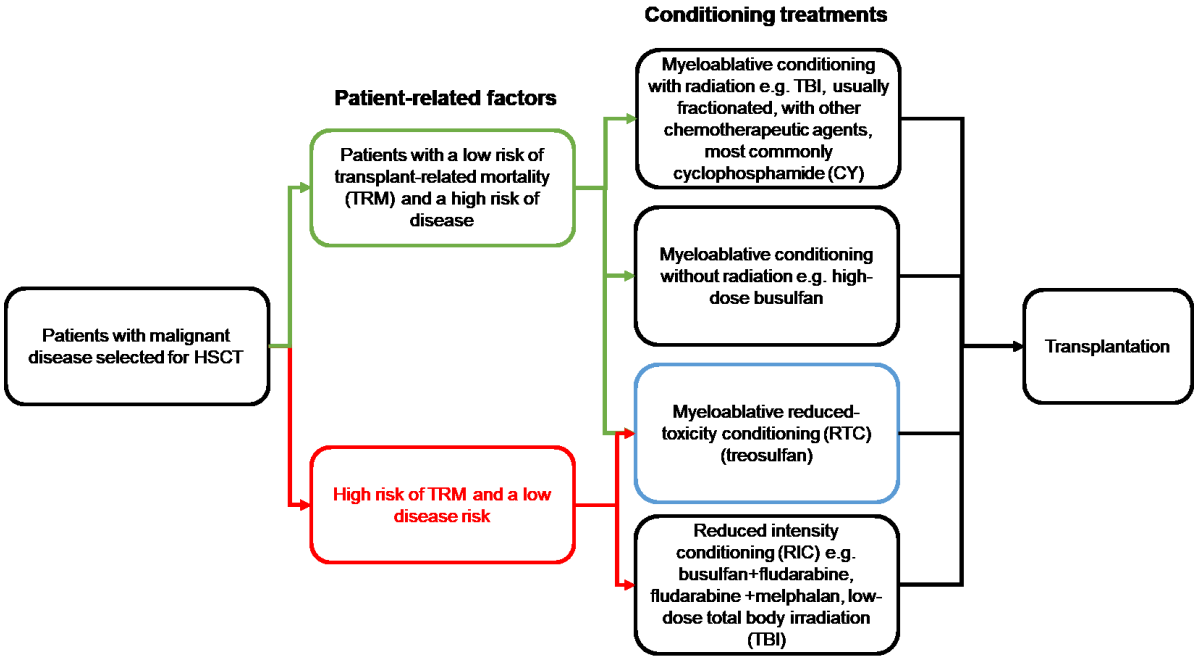
Abbreviations: ALL, acute lymphoblastic leukaemia; AML, acute myeloid leukaemia; BMTCTN, Blood and Marrow Transplant Clinical Trial Network; CMML, chronic myelomonocytic leukaemia; ELN, European Leukaemia Net; EBMT, European Society for Blood and Marrow Transplantation; HSCT, haematopoietic stem cell transplantation; IMFS, International Myelodysplastic Syndromes Foundation (IMSF); MDS, myelodysplastic syndrome; RFS, relapse-free survival; RIC, reduced-intensity conditioning. †The international expert panel was active within the EBMT the ELN, BMTCTN and IMSF.

In terms of HSCT timing, it is generally recommended that alloHSCT be pursued at early disease stages, possibly as soon as complete remission (CR) is achieved in patients with high-risk diseases.[8,19,24,26,27] This recommendation is driven by the fact that remission status at the time of HSCT is an important prognostic factor in predicting the risk of relapse in patients with both myeloid and lymphoid malignancies.[19,20,24]

B.1.3.2.1 Selection of conditioning therapy

The current treatment paradigm for conditioning therapies along with the proposed positioning of treosulfan is summarised in Figure 2.

Figure 2: Clinical context of treosulfan with fludarabine in conditioning treatment for malignant haematological diseases selected for HSCT



B.1.3.3 Conditioning therapy

Patients undergoing an alloHSCT are prepared with chemotherapy alone or chemotherapy combined with radiotherapy, the so-called conditioning or preparative regimen. This is the first time that NICE has evaluated a conditioning treatment for HSCT and it should be understood that a conditioning treatment is not a treatment for the underlying disease. Instead, there are three aims of conditioning treatment:

- to reduce the tumour burden when the disease is neoplastic,
- to eliminate the self-renewing capacity of the patient's own haematopoiesis, and
- to suppress the recipient's immune system in order to allow engraftment of stem cells.

Exceptions to this rule are infants with severe combined immune deficiency (SCID) and patients with severe aplastic anaemia (SAA) with an identical twin donor who may be grafted without conditioning.[28]

Transplant-related mortality (TRM) after myeloablative regimens increases with increasing patient age, and 50-55 years used to be considered an upper age limit.

With the aim of reducing toxicity and thus making transplantation available in the older patient population, so called non-myeloablative (NMA) conditioning regimens were developed.

By definition, myeloablative conditioning regimens (MAC) do not allow for autologous recovery and require stem cell support whereas NMA conditioning regimens do not require stem cell support. Regimens that do not match these criteria have been classified as reduced-intensity conditioning (RIC), whereby the dose of total body irradiation (TBI) or the alkylating agent is usually reduced by at least 30% compared with an ablative regimen. RIC regimens cause cytopenia of variable duration, and should be given with stem cell support, although cytopenia may not be irreversible.[29]

Although full consensus has not been reached in the HSCT community, conditioning regimens consists of three main types of regimens which are summarised in Table 5.

Table 5: Comparison of myeloablative conditioning, reduced-intensity conditioning and non-myeloablative regimens

	Myeloablative conditioning (MAC)	**Reduced intensity conditioning (RIC)**	**Non-myeloablative regimens**
Description	Total body irradiation (TBI)-based regimens and high-dose chemotherapy-based regimens.[24]	Regimens in which cytotoxic components of the conditioning regimen are reduced or replaced with less toxic but immunosuppressive agents.[24,30]	Typically consist of low-dose TBI alone and combination of fludarabine with low dose TBI or other agents.[24,31]
Examples	Traditionally alkylating agents such as high-dose busulfan and melphalan, are examples of chemotherapy-based regimens.[24]	Examples include fludarabine combined with lower doses of melphalan, busulfan or cyclophosphamide.[24,31]	Low-dose TBI plus fludarabine.[32]
Patient population	Generally younger, medically fit patients or those with no or minimal pre-existing comorbidities.[17]	Patients considered unsuitable for standard "classic" MAC because of comorbidities or age who would otherwise not be eligible for transplantation e.g. acute myeloid leukaemia (AML).	Allows options of HSCT for patients who were traditionally not eligible due to advanced age or comorbidities. Commonly used in lymphoproliferative diseases due to the indolent nature of some of these diseases.[33]
Summary	Causes irreversible cytopenia and stem cell support is required.[34]	Cytopenia of variable duration is induced which may not be irreversible.	Causes minimal cytopenia and can be given also

	Myeloablative conditioning (MAC)	Reduced intensity conditioning (RIC)	Non-myeloablative regimens
	Generally low relapse rates but morbidity, high treatment-related mortality and toxicity, multiple organ toxicity, and NRM restricts suitable patient population.[1,14,35–37] Prolonged time to engraftment and prolonged use of immunosuppression result in increased risk for bacterial, viral, and fungal infections.	Stem cell support is given.[34] Provides reduced dose and limited toxicity in order to lower NRM and is generally well tolerated. Reduced morbidity and transplant-related mortality (TRM). However, limitations include increased relapse rates.[38] Graft-versus-host disease (GvHD) needs improving.[21]	without stem cell support.[34] Considered by some as a sub-set of RIC regimens with the lowest conditioning doses of all. Post-transplant infection risk is reduced, however, there may be slower engraftment. Relies heavily on control of malignancy, termed graft-versus-tumour (GvT). GvT relies on the immune-mediated assistance from donor lymphocytes for complete eradication of malignant cells.[17]

B.1.3.3.1 Myeloablative conditioning regimens containing radiation

High dose total body irradiation (TBI) has been widely used as part of the conditioning regimen due to its stem cell toxicity, immunosuppressive properties, its effectiveness against most leukaemias and lymphomas, and its ability to penetrate to sanctuary sites (e.g. brain, spinal cord, testes). The majority of regimens combined 12-16 Gray (Gy) TBI, usually fractionated, with other chemotherapeutic agents, most commonly cyclophosphamide (CY), based on its antineoplastic and immuno-modulatory properties. In general, higher doses of TBI, although reducing the relapse risk, resulted in increased, often fatal gastrointestinal, hepatic, and pulmonary toxicities, secondary malignancies, and impaired growth and development in children.[39,40]

In addition to CY, various agents, such as cytarabine (AraC), etoposide (ETO), melphalan (MEL), and busulfan (BU), have been combined with TBI as conditioning regimens; however, due to the lack of randomised trials, there is currently no evidence suggesting that any of these combinations are superior to CY and high-dose TBI.[24,39]

B.1.3.3.2 Myeloablative conditioning regimens without radiation

Regimens have been developed in which TBI is replaced by additional chemotherapeutic agents. These approaches have primarily been developed for autologous transplantation, but they have also been used in the allogeneic setting. The primary advantage of regimens that lack TBI is reduced toxicity. Additionally, the cost is lower, the regimen is easier to administer, and radiation can still be given to sites of prior disease following transplantation. Alkylating agents remain the mainstay of such regimens, due to favourable toxicity profile (marrow toxicity as dose-limiting toxicity) and their effect on non-dividing tumour cells.[24]

Commonly used non-radiation-containing standard intensity conditioning regimens are frequently based on orally or intravenously administered BU and other cytotoxic agents. A regimen consisting of high-dose BU (16 mg/kg total dose) and CY (200 mg/kg total dose) was developed,[41] modified (total CY dose was decreased to 120 mg/kg),[42] and has been widely used since in patients undergoing autologous and allogeneic HSCT.

B.1.3.3.3 Nonmyeloablative (NMA) conditioning regimens

NMA regimens have been explored intensively during the last few years. Along with RIC regimens, NMA regimens have expanded the number of patients eligible for HSCT.

Due to lowered toxicity, NMA transplants can be appropriate for patients older than 55 years, which is a common upper limit for standard myeloablative transplantation as well as patients with one or more co-morbidities that would ordinarily exclude them from undergoing myeloablative transplantation.[43]

The purine analogue fludarabine (FLU) has been widely incorporated into such regimens. It is highly immunosuppressive, producing profound lymphopenia, which has been shown to facilitate allogeneic stem cell engraftment. It has the additional advantages of having anti-tumour activity in haematological malignancies and a low non-haematological toxicity profile.[44]

NMA regimens differ from RIC regimens in that the former may result in only minimal cytopenias that do not require stem cell support whereas RIC regimens do require

stem cell support.[29] Commonly used RIC conditioning regimens are shown in Table 6.

Table 6: Examples of NMA conditioning regimens

NMA regimens
FLU + CY + antithymocyte globulin (ATG)
FLU + AraC + idarubicin
Cladribine + AraC
Total lymphoid irradiation + ATG
TBI ≤ 2 Gy ± purine analogue

Source: Gyurkocza 2014[24]

It is important to consider that the myeloablative intensity, frequency, and combination of treatments in different conditioning regimens vary significantly; therefore, the distinction between truly myeloablative regimens and RIC regimens is not clearly defined.[8,9]

B.1.3.3.4 Reduced intensity conditioning (RIC) regimens

Myeloablative regimens are associated with considerable toxicity but the graft-versus-tumour (GvT) effects contributes to the eradication of malignant disorders. For patients with haematological malignancies, GvT effect is mediated by the allogeneic donor cells which requires the permanent engraftment of donor-type immunocompetent cells. However, it does not necessarily require a toxic myeloablative preparative regimen.

RIC were developed to make alloHSCT accessible to older patients and patients with pre-existing comorbidities who previously were not considered candidates for high-dose conditioning.[24,30]

Examples of RIC regimens are shown in Table 7. These RIC regimens require stem cell support.[29]

Table 7: Examples of RIC regimens

RIC regimens
TBI ≤ 500 cGy as a single fraction or ≤ 800 cGy if fractionated
Total BU dose ≤ 9 mg/kg
Total melphalan (MEL) dose < 140 mg/m²

RIC regimens
Thiotepa (TT) < 10 mg/kg

Source: Gyurkocza 2014[24]

While standard myeloablative regimens generally lead to low relapse rates, they are associated with high treatment-related toxicity and TRM.[6] Thus, patient who are not eligible for myeloablative regimens (e.g. elderly and patients with comorbidities) usually receive a RIC regimen to minimise treatment-related toxicity, NRM and TRM. Unfortunately, lower dose intensity usually comes along with a higher risk of relapse.[6]

B.1.3.3.5 Myeloablative reduced-toxicity conditioning (RTC) regimens

The so-called myeloablative reduced-toxicity conditioning (RTC) regimens such as treosulfan are designed to reduce treatment-related toxicity and relapse rates compared with conventional therapies.[8,9] RTC regimens aim to provide the efficacy of MAC regimens to patients who cannot tolerate such MAC regimens and would otherwise be treated with less efficacious RIC regimens.

B.1.3.4 Transplantation and engraftment

B.1.3.4.1 Transplantation source

The required haematopoietic stem cells can be collected from several sources, including bone marrow (BM), peripheral blood (PBSCT), and umbilical cord blood (UCB). Worldwide in 2012, grafts were collected from peripheral blood (66%), bone marrow (24%; mainly non-malignant disorders) and cord blood (10%).[45]

The donor sources include cells obtained from another person (termed allogeneic transplant), an identical twin (termed syngeneic transplant), or the patient itself (termed autologous transplant). Generally, candidates for autologous transplantation have no demonstrable malignancy in the blood or bone marrow. Whereas treatment-related morbidity and mortality rates are lowest with autografts, the major problem is tumour relapse. This finding relates to the absence of a GvT effect (i.e., immunologic attack on the tumour by immunocompetent T cells and natural killer cells in the donor graft) and the reinfusion of occult tumour cells in the graft.

Allogeneic transplants are associated with lower relapse rates compared with autologous transplants because of the GvT effect. However, transplant-related morbidity and mortality is considerable, mainly due to the development of infections as well as acute and chronic GvHD. Allogeneic stem cells may be obtained from the patients relatives (identical twins, human leukocyte antigen [HLA]-matched related donors [MRD], mismatched related donors [misMRD]) or unrelated donors (matched unrelated donors [MUD], umbilical cord blood [UCB]).

Although HLA-identical sibling donors remains the preferred donor source, survival after transplantation is comparable among patients receiving HSCTs from HLA-identical sibling and matched unrelated donors for several diseases.[10] The development of unrelated donor registries and increased utilisation of cord blood and partially matched related donor transplants have ensured a donor for essentially everyone who needs a transplant.[46] In 2012, more HSCTs were registered from unrelated donors (n = 16,433) than related donors (n =15,493).[45]

B.1.3.4.2 Engraftment

The rate at which stem cells are received and start to grow and make new cells is known as engraftment and is considered a step in successful transplantation. This engraftment is usually faster following PBSCT.[47]

B.1.3.5 Post-graft immunosuppression and post-transplant follow-up

In alloHSCT the donor cells provoke immunologic reactions involved in engraftment of the donor cells, graft-versus-host disease (GvHD), graft versus tumour (GvT) control of the malignancy, the development of tolerance, and immune reconstitution. These immunological reactions are monitored and in addition any infections must be treated. Immunological reactions are influenced by the conditioning regimen, the type and source of the donor graft as well as the post-graft immunosuppressive regimen which is utilised.[17]

The goal of post-graft immunosuppression is to obtain a delicate balance between improving clinical outcomes such as prevention of graft rejection and lowering the rate and/or severity of GvHD, whilst avoiding excessive immunosuppression.

Excessive immunosuppression is associated with more frequent and severe infectious complications including bacterial, fungal and viral infections which may contribute to non-relapse mortality.[17]

B.1.3.5.1 Complications of HSCT

There are a broad spectrum of acute and long-term complications, ranging from organ-specific complications and infections to secondary malignancies.

Prolonged, severe pancytopenia

Severe (< 500/μL but often < 100/μL) and prolonged (up to 4 weeks) neutropenia is common after transplantation and invariably requires the use of empiric broad-spectrum antibiotics and blood cell transfusions.

Graft rejection

Graft rejection means that donor cells fail to regenerate within the recipient. Mechanisms include the failure of immunosuppressive agents to inactivate the host immune system, inadequate ratio of donor cells to facilitator cells infused, drug injury to marrow, or viral infections.

Graft versus host disease

GvHD occurs when immunocompetent T cells and natural killer cells in the donor graft recognise host antigens as foreign targets and mediate a reaction. The disease may cause significant morbidity and mortality and has been divided into acute and chronic forms. To avoid the effects of acute GvHD, patients are given potent immunosuppressive drugs immediately before and for many months after transplantation. 50% to 70% of patients treated with alloHSCT develop chronic GvHD within ten years of treatment.[48] This syndrome, which resembles connective-tissue disorders such as scleroderma, may develop as a continuation of active acute GvHD (progressive), may occur after successful clearing of acute GvHD (quiescent), or may appear without antecedent acute GvHD (*de novo*). The target organs are more widespread than in acute GvHD. The onset of chronic GvHD is associated with fewer relapses, indicative of a GvT effect.[49]

Pulmonary complications (interstitial pneumonitis)

Interstitial pneumonitis is frequently a fatal syndrome often caused by viral infections. Lung injury can also be due to TBI or pulmonary toxins (carmustine).

Hepatic sinusoidal obstruction syndrome

Hepatic sinusoidal obstruction syndrome (HSOS), previously termed veno-occlusive disease (VOD) manifests as jaundice, tender hepatomegaly, unexplained weight gain, or ascites. The condition usually develops by 30 days after HSCT, although it can occur later. Historically, its reported incidence ranges from approximately 5 to 60%, and this variation is clearly not only related to the intensity of the conditioning regimen, the type of transplant and the presence of risk factors, but also on the clinical criteria used for HSOS diagnosis. Risk factors for this complication include a history of previous hepatocellular disease, certain conditioning regimens, advanced age, the presence of GvHD, the type of GvHD prophylaxis, poor performance status at transplantation, and the use of matched unrelated or mismatched donor grafts. Therapy for HSOS is generally unsatisfactory.[50]

Late-onset problems

Patients undergoing HSCT have an increased risk of secondary malignancy, most often occurring many years following the transplantation procedure. Transplant recipients are at higher risk of infection, including the reactivation of dormant herpes viruses. Gonadal dysfunction arises in up to 92% of men and up to 99% of women; its frequency depends on the timing of transplantation, on radiotherapy, and on other factors. The medications that transplant recipients need to take can impair liver function, and transfusion-associated haemosiderosis can do so as well. 40% to 50% of patients suffer from lipid metabolic disturbances that increase the risk of myocardial infarction, peripheral arterial occlusive disease, and stroke. Their life expectancy is shorter than that of the overall population.[48]

B.1.3.6 Guidelines and clinical practice

Although no consensus guidelines are available on the choice of optimal conditioning regimens there are guidelines on the indications for HSCT in a range of malignant diseases available from:

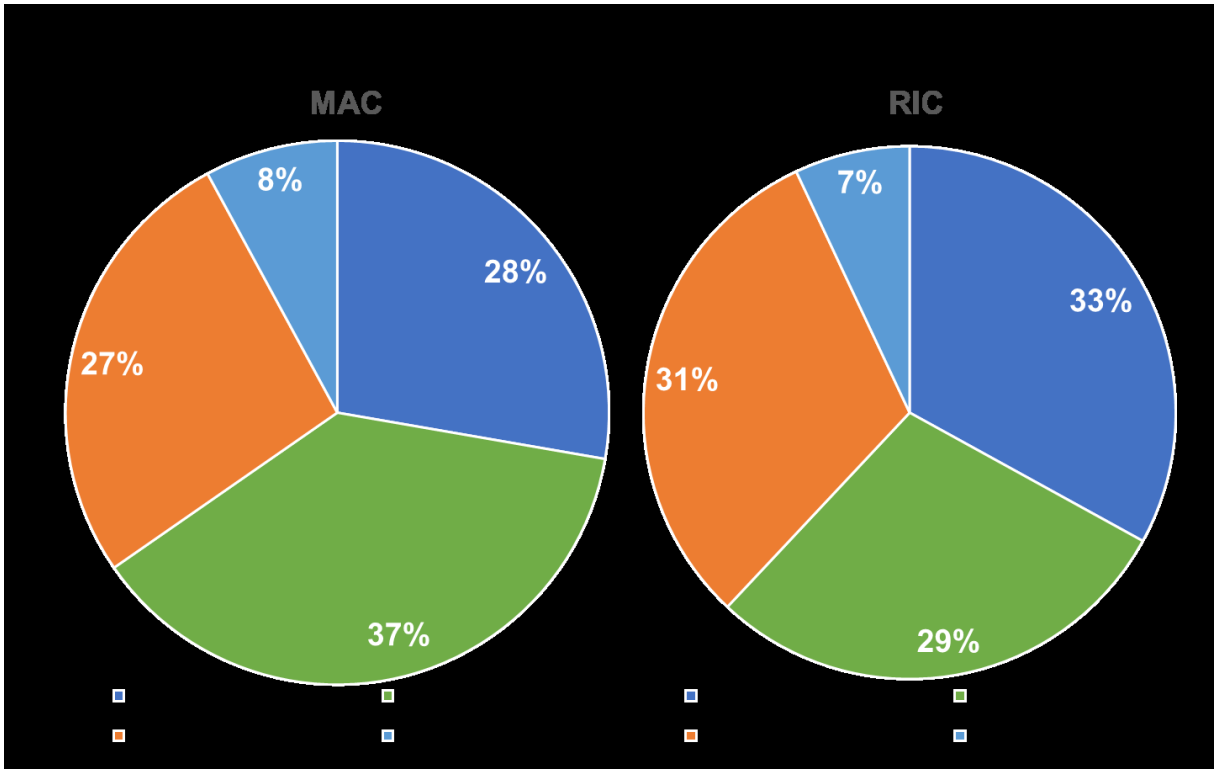
- The 2013 British Society of Blood and Marrow Transplantation (BSBMT) Transplantation Indications Table.[51]
- The 2017 international expert panel active within the EBMT the ELN, BMTCTN and IMSF.[20]
- The 2015 American Society for Blood and Marrow Transplantation American (ASBMT) clinical guidelines.[10]

The 2013 BSBMT and the 2015 ASBMT clinical guidelines provide guidance on "routine" indications for HSCT both within and external to the clinical trial setting.[10,51] According to both these clinical guidelines, auto- or alloHSCT is indicated in adults and children with the following malignant diseases:

- Malignant haematological disorders including leukaemia e.g. AML, ALL, CML, MDS and Hodgkin and non-Hodgkin lymphomas.
- Solid tumours (e.g. neuroblastoma, Wilms' tumour and sarcoma).

Standard conditioning regimens have not been established for the various indications; as a result, clinical practice varies among institutions. This is mainly due to the lack of comparative evidence from Phase III randomised controlled trials in the field.[24,52,53] The worldwide data (2007-20117) from the Center for International Blood and Marrow Transplant Research (CIBMTR) indicate that total body irradiation (TBI), combinations of busulfan with cyclophosphamide or fludarabine, and fludarabine combined with melphalan are among the most commonly used conditioning regimens for patients with AML or MDS undergoing myeloablative or RIC alloHSCT (Figure 3).[54]

Figure 3: Conditioning regimens commonly used in AML and MDS alloHSCT (CIBMTR data 2007-2017)



Source: DiSouza et al 2018.[54]
Abbreviations: allo, allogeneic; bu, busulfan; cy, cyclophosphamide; CIBMTR, Center for International Blood and Marrow Transplant Research; flu, fludarabine; HSCT, haematopoietic stem cell transplantation; MAC, myeloablative conditioning; Mel, melphalan; RIC, reduced-intensity conditioning; TBI, total body irradiation.

B.1.3.6.1 UK clinical care pathway

As with the rest of the world, in the UK the clinical care pathway for patients undergoing alloHSCT varies considerably depending on the underlying malignant disease, age and other co-morbidities. For this reason, NHS England has acknowledged the difficulty of developing national commissioning policy in this area as "previous attempts to develop evidence-based policies have highlighted the paucity of good quality evidence from randomised controlled trials, and the small size and poor quality of the studies upon which current clinical guidelines are based".[55]

In summary, the indications to perform either myeloablative or RIC conditioning HSCT are based on expert consensus and evidence-based reviews which are mainly derived from registry survey analyses and retrospective studies. Thus, there is limited direct evidence supporting the current practice in selecting patients with a wide range of malignancies (e.g. AML, ALL and MDS), for RIC versus MAC

regimens.[6] Therefore clinicians make decisions largely based on their own clinical assessment of the patient's individual characteristics.

B.1.3.7 Rationale for the development of treosulfan

Commonly used standard intensity conditioning regimens such as TBI/CY, BU/CY or combinations of BU with other cytotoxic agents are associated with a high risk of mortality and morbidity especially for intensively pre-treated patients. Therefore, these regimens are usually contraindicated in patients who are older (>55 years), have an increased risk for pulmonary or hepatic complications, have had organ or infectious complications with previous chemotherapy, as well as patients with previous autoHSCT. As a consequence, many efforts have been made to reduce the toxicity of these regimens while maintaining their efficacy with respect to donor cell engraftment and anti-tumour efficacy. The development of NMA and RIC regimens expanded the patient population that could receive alloHSCT; however, this was at the expense of a slightly increased relapse rate.

The ideal cytotoxic agent for conditioning regimens should have the following properties:

- sufficient stem cell toxicity (with respect to primitive as well as committed stem cells) and immunosuppression to enable rapid and stable engraftment,
- low organ toxicity, especially with respect to the liver, kidneys, lung, and the nervous system,
- sufficient cytotoxicity to guarantee an effective treatment of the underlying haematological malignancy,
- predictive pharmacokinetics (i.v. administration, linear pharmacokinetics, low inter-individual variability, no enzyme-dependent drug activation).

In the late 1990s, medac initiated a phase I clinical trial with high-dose chemotherapy followed by autoHSCT. The treosulfan dose could be escalated up to 5 times the conventional maximal tolerated dose (MTD) of 10 g/m². The MTD was reached at 47 g/m² administered as 2-hour infusions followed by autoHSCT, with diarrhoea, mucositis/ stomatitis, and metabolic acidosis being the dose-limiting toxicities. Severe liver toxicities such as HSOS, nephrotoxicities, neurotoxicities or lung

toxicities like those known to occur with other high-dose alkylating agents, including busulfan, were not observed after administration of high-dose treosulfan. [56]

From these encouraging results it was suggested that treosulfan could be an ideal drug for incorporation into a conditioning regimen prior to HSCT. Pilot studies have shown that treosulfan-based conditioning regimens followed by alloHSCT are feasible, tolerable, and effective and can be even used for patients who do not tolerate conventional standard intensity conditioning treatments.[57] These results endorsed the further clinical development of treosulfan in this setting. The primary goal was to establish a standard intensity but reduced toxicity conditioning regimen which in contrast to RIC regimens retains its myeloablative and antileukaemic activity thereby reducing the risk of relapse.

Prior to its development as a RTC agent, an oral formulation (and since 1990 an i.v. version) of treosulfan was marketed as Ovastat® and approved in several European countries as a treatment for patients with ovarian cancer. Over the decades many patients with ovarian cancer have therefore been treated with this agent and thus the toxicity profile of this medicinal product is well-known.

Compared to other cytotoxic agents treosulfan is very well tolerated. As a single agent, grade III/IV toxicities are very rarely observed. Myelosuppression is the most frequently observed side effect. Gastrointestinal toxicity, especially nausea/vomiting is rarely observed with treosulfan monotherapy and seldom exceeds grade II. Treosulfan lacks hepatic and renal toxicity and only very rarely causes lung toxicity. Due to its excellent tolerability, treosulfan is especially suited for patients with poor performance status or advanced age.[58]

B.1.4 Equality considerations

We do not envisage any equity or equality issues with the use of treosulfan in combination with fludarabine as conditioning treatment prior to alloHSCT in adults and children with malignant disease.

B.2 Clinical effectiveness

Clinical effectiveness summary

Summary of clinical evidence

- The efficacy and safety of the treosulfan-based conditioning therapy prior to alloHSCT in adult patients with various malignant haematological disorders was demonstrated in four prospective studies:
  - A large Phase III, prospective, randomised study (MC-FludT.14/L) comparing the efficacy and safety of treosulfan (10 g/m²/day) against busulfan (3.2 mg/kg/day), both in combination with fludarabine, in adult patients with AML or MDS (N=570) indicated for alloHSCT who were considered ineligible for standard conditioning therapy.[59,60]
  - Three Phase II prospective, non-randomised studies (MC-FludT.6/L[61],[62] MC FludT.7/AML[63] and MC-FludT.8/MDS[64]), investigating the efficacy and safety of high-dose treosulfan (10-14 g/m²/day) in combination with fludarabine in adult patients with various haematological malignancies indicated for alloHSCT (N=55, N=75 and N=46, respectively).[65–67]

Clinical effectiveness from MC-FludT.14/L Trial II

- MC-FludT.14/L Trial II is the pivotal phase III RCT assessing the safety and efficacy of treosulfan in combination with fludarabine for conditioning therapy as part of alloHSCT in adults with AML or MDS. With 570 randomised patients, it was the largest ever prospective RCT comparing two conditioning regimens.[60]
- The trial demonstrated that the treosulfan/fludarabine regimen was non-inferior to busulfan/fludarabine with regard to the primary endpoint (event-free survival (EFS) at 24 months after alloHSCT), and that the treatment difference for EFS was statistically significantly in favour of treosulfan:[60]
  - The EFS at 24 months after alloHSCT was 65.7% (95% confidence interval [CI]; 59.5, 71.2) for the treosulfan treatment group, and 51.2% (95% CI; 45.0, 57.0) for the busulfan treatment group, with a hazard ratio (HR) of 0.64 in favour of treosulfan (p=0.0000001).
  - The treatment difference for EFS between regimens was statistically significantly in favour of treosulfan (p=0.0005787), indicating a clinically relevant long-term advantage of treosulfan.
- The secondary endpoints assessed were: engraftment, complete donor-type chimerism, overall survival (OS), relapse/progression, NRM, TRM, GvHD-free survival (GRFS), Chronic GvHD-free and relapse/progression-free survival (CRFS), acute and chronic GvHD, adverse events (AEs), and Grade III/IV mucositis. The statistically significant advantage of the treosulfan/fludarabine regimen compared

with the busulfan/fludarabine regimen was demonstrated for the majority of the secondary endpoints:[60]

- The incidence of complete donor-type chimerism was statistically significantly higher with the treosulfan-based regimen both at day +28 (93.2% vs 83.3%; p=0.0159) and at day +100 (86.1% vs 80.2%; p=0.0381).
- The Kaplan-Meier estimate of OS at 24 months was significantly superior for patients treated with the treosulfan-based regimen than those treated with the busulfan-based regimen (72.7% vs 60.2%; p=0.0037).
- The cumulative incidence of NRM at 24 months after alloHSCT was statistically significantly lower in the treosulfan treatment group compared with the busulfan treatment group (12.0% vs 20.4%; p=0.0343).
- A statistically significant reduction in cumulative incidence of TRM at 24 months after alloHSCT was observed for the treosulfan-based regimen vs the busulfan-based regimen (12.8% vs 24.1%; p=0.0043).
- The rates of GRFS and CRFS were both statistically significantly higher in patients treated with the treosulfan-based regimen vs those treated with the busulfan-based regimen (GRFS, 50.3% vs 37.1%; p=0.0087; CRFS, 51.4% vs 37.2%; p=0.0030).
- Engraftment rates at day +28 were similar between the treosulfan-based regimen and the busulfan-based regimen (96.2% vs 96.8%; p=0.4235).
- Notably, the cumulative incidence of relapse/progression at 2 years after transplantation was comparable between the two conditioning regimens (treosulfan, 22.0%; busulfan, 25.2%; p=0.2631).

- The robustness of these results was confirmed by exploratory subgroup analyses (i.e. Kaplan Meier estimates of EFS by donor type, risk and age group, disease, HCT-CI score, and combinations thereof) and pre-planned sensitivity analyses (i.e. Cox regression models with different prognostic subgroups as factors or strata).[60]
- In the post-surveillance evaluation, follow-up data for EFS, OS, relapse/progression, and NRM showed a continued clinically relevant long-term advantage of treosulfan compared with busulfan.[60]

Clinical effectiveness from MC-FludT.6/L

- The phase II clinical study MC-FludT.6/L included 55 patients with various haematological malignancies (AML 35.0%, MDS 11.0%, MM 18.0%; CML 11.0, ALL 4.0, Mature B-cell neoplasms and other 20.0%, and Hodgkin's lymphoma 2.0%.) treated with treosulfan-based conditioning.[62]
- The study demonstrated treosulfan was tolerable and effective at all three treosulfan dose levels tested (10 g/m², 12 g/m² or 14 g/m²) and

produced sustained engraftment for granulocytes and leukocytes (98%) and for thrombocytes (80%). This sustained engraftment was confirmed by donor-type chimerism analysis.[62]

Clinical effectiveness from MC-FludT.7/AML

- Data from the phase II clinical study MC-FludT.7/AML in 75 AML patients treated with treosulfan conditioning confirmed the promising safety and efficacy of treosulfan-based conditioning therapy in AML patients.[63]
- By Day +28, engraftment was 93% for granulocyte, 100% for leukocyte, and 93% for platelets.[63]
- Severe (Common Terminology Criteria (CTC) grade III/IV) AEs were reported in 65% of the patients, but only 2 patients (3%) experienced a CTC grade IV event.[63]
- After a median follow-up of 715 days, the 2-year overall and disease-free survival (DFS) estimates were 61% and 55%, respectively. The 2-year incidences of relapse and non-relapse mortality reached 34% and 11%, respectively.[66]
- The study demonstrated that treosulfan-based conditioning regimen combine the high dose intensity of a standard MAC regimen with low non-haematological toxicities, comparable to RIC regimens, resulting in a low NRM without an obvious increase of the relapse risk.[63,66]

Clinical effectiveness from MC-FludT.8/MDS

- Data from the phase II clinical study MC-FludT.8/MDS in 46 MDS patients treated with treosulfan conditioning confirmed the low acute non-haematological toxicity of the regimen and support the use of the term “reduced-toxicity conditioning” for treosulfan.[64]
- The conditional cumulative incidences of granulocyte, leukocyte, and platelet engraftment by Day +28 were 96%, 96%, and 87%, respectively.[64]
- Severe (CTC grade III/IV) AEs were reported in 87% of the patients, but only 4 patients (9%) experienced in total 6 episodes of CTC grade IV AEs.[64]
- Acute and chronic GvHD incidences and severity were in the range of those commonly reported for standard conditioning regimens in alloHSCT. The resulting NRM rates are low and encouraging with 9% NRM at 100 days and 17% NRM at 720 days appear similar to those seen after RIC transplantation and lower than those reported after standard conditioning in MDS patients, in whom 100-day rates between 19% and 27% have been observed.[64,67]

Paediatric clinical effectiveness phase II trial MC-FludT.17/M)

- Long-term follow up for the MC-FludT.17/M trial[63] is ongoing but the primary endpoint, the rate for freedom from TRM until 100 days after HSCT, was 98.6% (90% CI: 93.4, 99.9) indicating the promising tolerability and safety of this regimen. The safety analysis along with the Kaplan-Meier estimation of NRM at 12 months supports this observation.
- Based on the data approaching 100% engraftment and >90%, for chimerism data, efficacy parameters like EFS, OS, and GvHD-free and relapse-free survival confirm the usefulness of this alternative conditioning treatment.
- In the selected paediatric population, treosulfan has demonstrated a positive benefit-risk profile.

B.2.1 Identification and selection of relevant studies

See appendix D for full details of the process and methods used to identify and select the clinical evidence relevant to the technology being appraised.

To assess the comparative efficacy and safety of treosulfan with fludarabine as a conditioning treatment for malignant diseases prior to alloHSCT in paediatric and adult patients, a systematic literature review (SLR) was conducted to identify randomised control trials (RCTs) and non-RCTs.

Full details of the methodology used to identify and select the RCT and non-RCT clinical evidence relevant to the technology being appraised are reported in Appendix D.1.1. A full summary of the included and excluded studies, including the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram and reasons for exclusion, are also provided in Appendix D.1.

B.2.2 List of relevant clinical effectiveness evidence

B.2.2.1 Pivotal treosulfan effectiveness study

The clinical effectiveness evidence presented in this submission is focused on MC-FludT.14/L Trial II,[60] the pivotal phase III RCT assessing the safety and efficacy of treosulfan in combination with fludarabine for conditioning therapy as part of

alloHSCT in adults with AML or MDS (see Table 8). With a total of 570 randomised patients, this study is the largest prospective RCT comparing two conditioning regimens ever performed.

Table 8: Clinical effectiveness evidence - MC-FludT.14/L Trial II

Study	**MC-FludT.14/L Trial II**				
Study design	Randomised, parallel-group, open label, multicentre, international, group-sequential Phase III non-inferiority trial				
Population	Adult patients with AML or MDS with increased risk for standard conditioning therapies (age ≥50 years and/or HCT-co-morbidity index [HCT-CI] score >2). All included patients were not eligible for a standard MAC busulfan- or TBI-based regimen				
Intervention(s)	Treosulfan i.v. (10 g/m²/day [d-4 to d-2]) + fludarabine i.v. (30 mg/m²/day [d -6 to d -2]) prior to alloHSCT				
Comparator(s)	Dose-reduced busulfan plus fludarabine: Busulfan i.v. (3.2 mg/kg/day [d-4 to d-3]) + fludarabine i.v. (30 mg/m²/day [d-6 to d-2]) prior to alloHSCT				
Indicate if trial supports application for marketing authorisation	Yes	✓	Indicate if trial used in the economic model	Yes	✓
	No			No	
Rationale for use/non-use in the model	MC-FludT.14/L is the pivotal trial for treosulfan.				
Reported outcomes specified in the decision problem	• **overall survival (OS)** • **event-free survival (EFS)** • **rates of relapse** • success of stem cell transplantation (engraftment) • **adverse effects of treatment**				
All other reported outcomes	• NRM				

Source: MC-FludT.14/L Trial II[60]
Outcomes in bold are incorporated in the economic model.

B.2.2.2 Further relevant treosulfan studies

Three phase II prospective, non-randomised studies sponsored by medac, (MC-FludT.6/L, MC-FludT.7/AML and MC-FludT.8/MDS), investigated the efficacy and safety of high-dose treosulfan (10-14 g/m²/day [d-6 to d-4]) in combination with fludarabine (30 mg/m²/day [d-6 to d-2]) followed by alloHSCT in adult patients with various haematological malignancies. These three studies have not been included in the economic model because they were not comparator studies and they had smaller patient numbers, therefore, the model was based on the randomized phase III data from the pivotal MC-FludT.14/L Trial II.

Study design, patient population and efficacy results from the three studies are summarised in Table 9.

Table 9: Summary of further relevant phase II treosulfan studies

Trial number (acronym)	MC-FludT.6/L	MC-FludT.7/AML	MC-FludT.8/MDS
Location	Finland, Germany, Poland and Sweden	Finland, Germany, Poland, and Sweden	Finland, Germany, Poland, and Sweden
Trial design	Phase II prospective, non-randomised studies	Phase II prospective, non-randomised studies	Phase II prospective, non-randomised studies
Number of patients	N=55	N=75	N=46
Eligibility criteria for participants	Adult patients with a haematological chemosensitive malignancy indicated for an alloHSCT, but presenting an increased toxicity risk for classical (high-dose busulfan or standard-dose total body irradiation) conditioning therapies. Malignancies included AML 35.0%, MDS 11.0%, MM 18.0%; CML 11.0, ALL 4.0, Mature B-cell neoplasms and other 20.0%, and Hodgkin's lymphoma 2.0%.	Adult patients with AML according to World Health Organization (WHO) classification (> 20% myeloblasts in peripheral blood or bone marrow at initial diagnosis) with < 5% myeloblast in the bone marrow, indicated for alloHSCT.	Adult patients with MDS according to WHO classification (< 20 % myeloblasts in peripheral blood or bone marrow at initial diagnosis) indicated for alloHSCT.
Settings and locations where the data were collected	Specialist HSCT centres	Specialist HSCT centres	Specialist HSCT centres
Trial drugs	10 g/m², 12 g/m² or 14 g/m² i.v. infusion of treosulfan, day -6, -5, -4.	14 g/m²/d i.v. infusion of treosulfan, day -6 to -4.	14 g/m²/d i.v. infusion of treosulfan, day -6 to -4.
Primary objectives	• Evaluation of feasibility and tolerability of 3 x 10, 12 or 14 g/m² treosulfan combined with 5 x 30 mg/m² fludarabine prior to alloHSCT. • Documentation of engraftment, leuko- and thrombopoiesis. • Quantification of donor-type chimerism. Time Frame: 6 months	Efficacy: • Evaluation of engraftment. Safety • Evaluation of the incidence of the following CTC grade III and IV adverse events between day -6 and day +28: - Hyperbilirubinemia and mucositis / stomatitis	Efficacy: • Evaluation of engraftment. Safety: • Evaluation of the incidence of the following CTC grade III and IV adverse events between day -6 and day +28: - Hyperbilirubinemia and mucositis / stomatitis

Trial number (acronym)	MC-FludT.6/L	MC-FludT.7/AML	MC-FludT.8/MDS
		- Veno-occlusive disease - Seizures. Time Frame:3.5 years	- Veno-occlusive disease - Seizures. Time Frame: 4 years
Summary of primary endpoint results	• Treosulfan conditioning treatment was tolerable and effective at all three treosulfan dose levels tested. • Sustained engraftment 98% for granulocytes and leukocytes and 80% for thrombocytes. • Sustained engraftment was confirmed by donor-type chimerism analysis in bone marrow or peripheral blood.	• The conditional cumulative incidences of granulocyte, leukocyte, and platelet engraftment by Day +28 were 93%, 100%, and 93%, respectively. • Severe (CTC grade III/IV) AEs were reported in 65% of the patients, but only 2 patients (3%) experienced a CTC grade IV event.	• The conditional cumulative incidences of granulocyte, leukocyte, and platelet engraftment by Day +28 were 96%, 96%, and 87%, respectively. • Severe (CTC grade III/IV) AEs were reported in 87% of the patients, but only 4 patients (9%) experienced in total 6 episodes of CTC grade IV AEs.
Secondary objectives	• Evaluation of frequency and severity of acute and chronic graft-versus-host disease (GvHD) until 6 months after transplantation. • Evaluation of the proportion of relapse- and/or progression-free patients 6 months after transplantation (using standard remission criteria). • Evaluation of the proportion of surviving patients 6 months after transplantation. Time Frame: 6 months	Efficacy • Evaluation of disease free survival (DFS), overall survival (OS) and relapse incidence (RI) during total follow-up. • Donor chimerism on day +28, +56 and +100. Safety • Evaluation of incidence of non-relapse mortality (NRM) on Day +28 and Day +100 and the cumulative incidence of NRM during total follow-up. • Evaluation of the incidence of CTC grade III and IV AEs between Day -6 and Day +28. • Evaluation of cumulative incidence and severity of acute GvHD (until Day	Efficacy • Evaluation of disease free survival (DFS), overall survival (OS) and relapse incidence (RI) during total follow-up. • Donor chimerism on Days +28, +56 and +100. • Safety • Evaluation of incidence of NRM on Day +28 and Day +100 and the cumulative incidence of NRM during total follow up. • Evaluation of the incidence of CTC grade III and IV adverse events between Day -6 and Day +28. • Evaluation of cumulative incidence and severity of acute GvHD (until Day

Trial number (acronym)	MC-FludT.6/L	MC-FludT.7/AML	MC-FludT.8/MDS
		+100) and chronic GvHD during total follow-up. Time Frame: 3.5 years	+100) and chronic GvHD during total follow-up. Time Frame: 4 years
Summary of secondary endpoints results	• Cumulative incidence of acute GvHD of 65% (overall grade I – IV) and chronic GvHD of 57 % (limited and extensive), respectively. • Overall a cumulative incidence of relapse of 31 % after 2 years of follow-up of surviving patients was calculated. • At 6 months cumulative incidence was more or less identical in the three treosulfan dose groups, resulting in 20% (95 % CI: 9%, 31%) relapse or progression in the overall patient population. • Relapse incidence increased in all dose groups and reached 31% (95 % CI: 19 %, 44 %) in the overall study population 24 months after transplantation.	• Kaplan-Meier estimates by Day +720 were 55% for DFS and 61% for OS. • Cumulative incidence of relapse by Day +100 was low and increased to 34% by Day +720. • By Day +100 visit, the cumulative incidence rate of complete donor type chimerism increased to 92 % for all patients. • No case of non-relapse mortality (NRM) had occurred by Day +28. The cumulative incidence of NRM was 3% by day +100 and reached 11% by Day +720. • Severe (CTC grade III/IV) AEs were reported in 65% of the patients, but only 2 patients (3%) experienced a CTC grade IV event. • Acute GvHD (grade III-IV) occurred with a cumulative incidence of 11 % by Day +100, and extensive chronic GvHD occurred with a cumulative incidence of 16 % by Day +720.	• Kaplan-Meier estimates by Day +720 were 67% for DFS and 71% for OS. • Cumulative incidence of relapse was low and reached 16% by Day +720. • Sustained engraftment was confirmed by donor-type chimerism analysis in the bone marrow which reached 93% by Day +56 and Day +100 visit. • No case of NRM had occurred by Day +28. After a median follow-up of 780 days the cumulative incidences of NRM for all patients were 9% by Day +100, 13% by Day +360 and 17% by Day +720. • Severe (CTC grade III/IV) AEs were reported in 87% of the patients, but only 4 patients (9%) experienced in total 6 episodes of CTC grade IV AEs. • The cumulative incidence of acute GvHD by Day +100 was 56% for all grades and 16% for grades III/IV. • The estimated cumulative incidences of chronic GvHD were 59% for all grades, and 28% for extensive chronic GvHD.

Source: MC-FludT.6/L[62,65]; MC-FludT.7/AML[63,66]; and MC-FludT.8/MDS[64,67]

Engraftment: In the three studies (MC FludT.6/L; MC FludT.7/AML; and MC-FludT.8/MDS), the treosulfan/fludarabine conditioning regimen was demonstrated to enable a rapid and sustained engraftment of neutrophils (93-98% of patients), leukocytes (96-100% of patients) and platelets (80-93% of patients) by day +28 after alloHSCT. The observed median time to engraftment across the three studies ranged between 14 days and 20 days. The sustained engraftment was confirmed by a high incidence (92-94%) of complete donor-type chimerism at 100 days after transplantation across the three studies.

Relapse: The cumulative incidence of relapse across the three studies ranged between 4% and 11 % at day +100, between 16% and 30% at 12 months, and between 16% and 34% at 24 months. In study MC-FludT.6/L, the incidence of relapse across the three dose groups (10 g/m²/day, 12 g/m²/day and 14 g/m²/day [d 6 to d 4] treosulfan) showed a trend in favour of the 14 g/m²/day dose group vs the 10 g/m^2/day dose group (24% [95% CI; 4%, 43%] vs 40% [95% CI; 18%, 62%]).

NRM: Across the three studies, the observed cumulative incidence of NRM at 24 months after conditioning with treosulfan/fludarabine was between 11% and 20%. The main reasons for NRM were infections and acute or chronic GvHD. A difference between studies was observed in the NRM rates at 24 months for the treosulfan dose of 14 g/m^2/day [d 6 to d 4] (11% in MC FludT.7/AML; 17% in MC-FludT.8/MDS; 20% in MC FludT.6/L). These differences may be explained by a more homogenous patient population in studies MC-FludT.7/AML and MC-FludT.8/MDS (mainly patients with AML or MDS in CR), and the lower proportion of patients considered ineligible for standard conditioning compared with study MC FludT.6/L (high risk patients: 17% and 18% vs 98%). The Kaplan-Meier estimates of progression-free survival (PFS) and OS at 24 months after conditioning with treosulfan/fludarabine and alloHSCT ranged between 49-67 % and 61-71%, respectively.

Efficacy conclusions: Results from studies MC-FludT.6/L, MC-FludT.7/AML and MC-FludT.8/MDS show that treosulfan in combination with fludarabine is a highly efficacious conditioning therapy prior to alloHSCT for patients with various haematological malignancies. The efficacy parameters measured (engraftment, establishment of complete donor-type chimerism, incidence of relapse, NRM, PFS

and OS) reached very acceptable values irrespective of the study populations analysed (e.g. high-risk patients or patients in CR). Importantly, the observed cumulative NRM rates were relatively low, even in patient populations considered ineligible for standard MAC therapies.

B.2.2.3 Further ongoing paediatric studies

Study MC-FludT.16/NM is an ongoing Phase II, prospective, randomised, multicentre, open-label, active-controlled, parallel-group study comparing the efficacy and safety of the treosulfan/fludarabine regimen vs a busulfan/fludarabine regimen, both with or without thiotepa, in children (N=100) with non-malignant diseases indicated for first myeloablative alloHSCT (see Table 10).[68] The duration of the study to document all data required for final evaluation of the observation phase and follow-up phases is approximately 33 months. Final results from this study are expected in 2021.

MC-FludT.17/M was a Phase II, prospective, single arm, open-label, multicentre, non-controlled study designed to assess safety and efficacy of treosulfan as part of a standardised fludarabine-containing conditioning therapy prior to alloHSCT and to contribute to a pharmacokinetics model which would permit, in conjunction with data from the clinical trial MC-FludT.14/L, to extend the use of treosulfan to the paediatric population by extrapolating efficacy and safety results. The study enrolled paediatric patients aged 28 days to <18 years with haematological malignancies (i.e. ALL, AML, MDS, or juvenile myelomonocytic leukaemias [JMML]), indicated for alloHSCT (see Table 10). The study is not yet completed as long-term (3 years) follow-up is on-going for some patients.

Table 10: Summary of further on-going paediatric treosulfan studies

Trial number (acronym)	**MC-FludT.16/NM**	**MC-FludT.17/M**
Location	Czech Republic, Germany, Italy, Poland	Czech Republic, Germany, Italy, Poland, United Kingdom
Trial design	An ongoing Phase II, prospective, randomised, multicentre, open-label, active-controlled, parallel-group study.	A Phase II, prospective, single arm, open-label, multicentre, non-controlled study.
Number of patients	N=100	N=70
Eligibility criteria for participants	Children up to 17 years with non-malignant disease indicated for first myeloablative allogeneic HSCT, including inborn errors of metabolism, primary immunodeficiencies, haemoglobinopathies and bone marrow failure syndromes.	Children up to 17 years with Haematological malignant disease i.e. ALL, AML, MDS or JMML, indicated for alloHSCT.
Settings and locations where the data were collected	Specialist HSCT centres	Specialist HSCT centres
Trial drugs	Treosulfan: 10, 12 or 14 g/m^2/day, based on body surface area, administered i.v. on three consecutive days (-6, -5 and -4); Busulfan: 3.2 to 4.8 mg/kg/day, based on body weight) according to authorised dosage for children and adolescents administered i.v. as part of the background conditioning regimen on four consecutive days (days -7, -6, -5 and -4); given in 1, 2, or 4 portions per day according to the respective hospital's standard.	Treosulfan: i.v., BSA adapted: 10, 12 or 14 g/m²/day within 120 min to be administered prior to fludarabine; Fludarabine: i.v., 30 mg/m2/day on days from -7 to -3 prior to HSCT. With or without Thiotepa: i.v., 2 x 5mg/kg/day on day -2 depending on investigator choice.
Primary objectives	Comparative evaluation of freedom from transplant (treatment)-related mortality TRM, defined as death from any transplant-related cause from the day of first administration of study medication (day -7) until day +100 after HSCT. Time Frame: day -7 to day +100.	Freedom from transplant (treatment)-related mortality (TRM). Time Frame: from the day of first administration of study medication until day +100 after HSCT.

Trial number (acronym)	MC-FludT.16/NM	MC-FludT.17/M
Primary endpoint results	Trial on-going. Results not yet mature.	The rate for freedom from transplant (treatment)-related mortality until 100 days after HSCT was [redacted]% (90% CI: [redacted], [redacted]).
Secondary outcomes	Comparative evaluation of engraftment after HSCT Comparative evaluation of safety including early toxicity Comparative evaluation of hepatic sinusoidal obstruction syndrome (HSOS), lung toxicity (common Terminology Criteria for Adverse Events [CTCAE] term pulmonary fibrosis), hepatic toxicity (according Bearman's criteria) and infections of any CTCAE grade (non-serious and serious) until day +100. Comparative evaluation of donor-type chimerism on day +28, day +100 and 12 months after HSCT Comparative evaluation of overall survival (OS) until 12 months after HSCT. Comparative evaluation of primary and secondary graft failure until 12 months after HSCT. Comparative evaluation of incidence and severity of acute GvHD [aGvHD] (until day +100) and chronic GvHD [cGvHD] (until 12 months after HSCT). Comparative evaluation of use of rescue therapies including donor-lymphocyte infusions (DLIs), stem cell infusions with or without further conditioning regimens, re-occurrence of transfusion dependence (i.e. necessity of regular transfusions of red blood cells or platelets. Evaluation of PK parameters of treosulfan and its epoxides and to develop a PK model for assessing relevant covariates. Comparative evaluation of secondary graft failure, cGvHD, donor-type chimerism, OS and TRM during the longer-term follow-up phase.	Engraftment after HSCT [Time Frame: until engraftment] Safety including early toxicity until day +100 after HSCT, serious adverse reactions (SARs) until the end of the longer-term follow-up phase [Time Frame: until 12 months after HSCT] Hepatic sinusoidal obstruction syndrome (HSOS), lung toxicity (CTCAE term pulmonary fibrosis), hepatic toxicity and infections of any CTCAE grade (non-serious and serious) [Time Frame: until day +100 after HSCT] Donor-type chimerism [Time Frame: on day +28, day +100 and 12 months after HSCT] NRM, TRM, graft failure rate, incidence of relapse/progression, relapse-free/progression-free survival (RFS/PFS) and OS [Time Frame: after 12 months after HSCT and until the end of the longer-term follow-up phase] Incidence and severity of acute (until day +100) and chronic (until 12 months after HSCT) graft versus host disease (aGvHD/cGvHD) [Time Frame: until 12 months after HSCT] Use of rescue therapies including donor-lymphocyte infusions (DLIs) and further conditioning regimens [Time Frame: until 12 months after HSCT] PK parameters of treosulfan and its epoxides [Time Frame: day -6 prior to HSCT]

Trial number (acronym)	MC-FludT.16/NM	MC-FludT.17/M
Secondary endpoint results	Trial ongoing. Results not yet mature.	As of analysis of 12th March 2018, the key results of the further endpoints were: [redacted] subjects ([redacted]%) had died from a transplant-related cause. The Kaplan-Meier estimate of TRM was [redacted]% (90% CI: [redacted], [redacted]) at 12 months. [redacted] subjects ([redacted]%) were alive. The Kaplan-Meier estimate of OS at 12 months after HSCT was [redacted]% (90% CI: [redacted], [redacted]). [redacted] subjects ([redacted]%) had experienced disease relapse / progression. The cumulative incidence of relapse / progression at 12 months was [redacted]% (90% CI: [redacted], [redacted]). [redacted] subjects ([redacted]%) had died without relapse. The cumulative incidence of NRM at 12 months was [redacted] (90% CI: [redacted], [redacted]). [redacted] subjects ([redacted]%) had experienced disease relapse / progression and 2 subjects had died without previous relapse / progression. The resulting Kaplan-Meier estimate of relapse / PFS at 12 months was [redacted]% (90% CI: [redacted], [redacted]). No subjects experienced a primary graft failure, and [redacted] subject ([redacted]%) experienced a secondary graft failure. The number of subjects with reconstitution of granulopoiesis was [redacted] ([redacted]%). The maximum conditional cumulative incidence reached was [redacted]% (90% CI: 97.7, 100.0). All subjects experienced neutropenia that had a median duration of 22 days. The number of subjects with reconstitution of leukopoiesis was [redacted] ([redacted]%). The maximum conditional cumulative incidence reached was [redacted]% (90% CI: [redacted], [redacted]). All subjects experienced leucopenia that had a median duration of 20 days.

Trial number (acronym)	MC-FludT.16/NM	MC-FludT.17/M
		The number of subjects with reconstitution of thrombopoiesis > 20 x 10^9/L was [redacted] ([redacted]%), and the number of subjects with reconstitution of thrombopoiesis > 50 x 10^9/L was [redacted] ([redacted]%). The maximum conditional cumulative incidence reached was [redacted]% (90% CI: [redacted], [redacted]) for reconstitution of thrombopoiesis > 20 x 10^9/L and [redacted]% (90% CI: [redacted], [redacted]) reconstitution of thrombopoiesis > 50 x 10^9/L. At visit Day +28, the incidence of complete donor-type chimerism was [redacted]% (90% CI: [redacted], [redacted]), at visit Day +100 the incidence was [redacted]% (90% CI: [redacted], [redacted]), and at visit Month 12 the incidence was [redacted]% (90% CI: [redacted]*, [redacted]). [redacted] subjects ([redacted]) had experienced relapse / progression, graft failure or death. The Kaplan-Meier estimate of EFS at 12 months was [redacted]% (90% CI: [redacted], [redacted]). A total of [redacted] subjects ([redacted]%) were alive and had not experienced aGvHD > Grade III or moderate or severe cGvHD or relapse / progression. The Kaplan-Meier estimate of GvHD-free and relapse / progression-free survival (GRFS) at 12 months was [redacted]% (90% CI: [redacted], [redacted]). A total of [redacted] subjects ([redacted]%) were alive and had not experienced moderate or severe cGvHD or relapse / progression. The Kaplan-Meier estimate of chronic GvHD-free and relapse / progression-free survival (CRFS) at 12 months was [redacted]% (90% CI: [redacted], [redacted]). [redacted] subjects ([redacted]%) had used any rescue therapies.
Estimated study completion date	December 2019	Last subject completes long-term follow-up in September 2019

Source: MC-FludT.16/NM[68] and MC-FludT.17/M[63]

The MC-FludT.17/M trial is ongoing (with long-term follow up continuing until September 2019). However the primary endpoint of this trial, the rate for freedom from transplant (treatment)-related mortality until 100 days after HSCT, was [redacted]% (90% CI: [redacted], [redacted]) indicating the promising tolerability and safety of this regimen. The safety analysis along with the Kaplan-Meier estimation of NRM at 12 months supports this observation.[63]

Based on the engraftment and chimerism data approaching [redacted]% and >[redacted]%, respectively, efficacy parameters like EFS, OS, and GvHD-free and relapse-free survival confirm the usefulness of this alternative conditioning treatment.[63]

Overall, the reported safety and efficacy results of this Phase 2 alloHSCT trial demonstrated a positive benefit-risk for the treosulfan-based conditioning regimen used in the selected paediatric population and thus allowing extension of the use of treosulfan to this paediatric population.

B.2.2.4 Analysis of the European Society for Blood and Marrow Transplantation (EBMT) Registry data

The EBMT Registry contains patient clinical data, including aspects of the diagnosis and disease, first-line treatments, HSCT- or cell-therapy-associated procedures, transplant type, donor type, stem cell source, complications and outcome. As of 2018, the EBMT Registry has data on over half a million patients given a HSCT procedure in Europe.[69]

Medac commissioned the EBMT to conduct a Registry-based study to re-analyse partly published EBMT registry data on fludarabine/melphalan (Flu/Mel) and busulfan/cyclophosphamide (BU/CY) based conditioning treatment without TBI (Control EBMT) compared to fludarabine/treosulfan based conditioning of adult AML and MDS patients treated in MC-FludT.14/L phase III (Treated chemotherapy [CT]) by using propensity score matching methods.

Table 11: Summary of EBMT Registry Analysis

Trial number (acronym)	EBMT Registry Analysis
Location	Europe
Trial design	Medac commissioned registry-based retrospective study of partly published EBMT registry data base using propensity score matching methods
Number of patients	Control EBMT - [redacted] patients Treated CT - [redacted] patients
Eligibility criteria for participants	• Patients from European Union countries aged between 50 and 70 with primary or secondary AML in CR or MDS patients (regardless of disease subtype (WHO) stage) • Donor type HLA identical sibling/MRD or MUD with source of stem cells being bone marrow or peripheral blood • Patients undergoing first alloHSCT between 2010 and 2016 with a Karnofsky score >=60 and utilising Flu/Mel or BU/CY without TBI conditioning
Settings and locations where the data were collected	EBMT Registry database (patients from Registry from 2010 – 2016)
Trial drugs	Control EBMT - fludarabine/melphalan (Flu/Mel) and busulfan/cyclophosphamide (BU/CY) based conditioning treatment without TBI Treated CT - fludarabine / treosulfan based conditioning of adult AML and MDS patients treated in MC-FludT.14/L Trial II
Primary objectives	Key study objectives were to: • Compare OS at two years after alloHSCT of fludarabine/melphalan and busulfan/cyclophosphamide based conditioning treatment with fludarabine/treosulfan based conditioning in MDS and AML patients separately. • Compare cumulative incidence of relapse (RI) at two years after transplant of fludarabine/melphalan and busulfan/cyclophosphamide based conditioning treatment with fludarabine/treosulfan based conditioning in MDS and AML patients separately. • Compare NRM at two years after transplant of fludarabine/melphalan and busulfan/cyclophosphamide based conditioning treatment with fludarabine/treosulfan based conditioning in MDS and AML patients separately.
Summary of primary endpoint results	• For MDS patients, treatment with fludarabine/treosulfan based conditioning leads to improved OS vs BuCy (HR [redacted] (95% CI [redacted]- [redacted]), p=[redacted]). Moreover, other comparisons (vs fludarabine + melphalan [FluMel]) were consistently in favour of fludarabine/treosulfan (HR < [redacted]). The differences were clinically relevant (over [redacted]% difference for OS, up to [redacted]% difference for NRM, [redacted]% difference for RI at 2 years). • For AML patients, OS (HR [redacted] (95% CI [redacted] - [redacted]), p<[redacted]) vs FluMel; HR [redacted] (95% CI [redacted]- [redacted]), p=[redacted] vs BuCy] is significantly improved, while NRM is significantly reduced (HR [redacted] (95% CI [redacted]- [redacted]), p=[redacted]) vs FluMel; HR [redacted] (95% CI [redacted] - [redacted]), p= [redacted] vs BuCy) with the use of fludarabine/treosulfan based conditioning compared to FluMel and BuCy.

Source: EBMT Report 2019[69]

In the Registry analyses, care was taken to select the most appropriate propensity score matching method and the resultant matched cohorts are overall well balanced, providing the optimal estimates of treatment effect given the data. Compared to BuCy and FluMel, using treosulfan-based conditioning leads to significantly improved OS and significantly lower NRM while leaving relapse incidences relatively unaffected. Sensitivity analyses by means of Cox regression analyses clearly support these results.

Therefore a significant clinical benefit of the treosulfan-based conditioning from the MC-FludT.14/L Trial II was demonstrated in comparison to cyclophosphamide (BuCy), as well as melphalan (FluMel)- based conditioning regimens in MDS (OS p=[redacted] vs BuCy) as well as AML patients (OS p<[redacted] vs FluMel and p=[redacted] vs BuCY).[69]

B.2.2.5 Definition of endpoints utilized in MC-FludT.14/L Trial II

There follows short definitions of the endpoints utilised in the pivotal MC-FludT.14/L Trial II. Further details can be found in Appendix L.

EFS: Event-free survival within 2 years after transplantation was defined as the primary endpoint of the trial. EFS was measured from time of end of HSCT (= Day 0) to time of event. Events were defined as relapse of disease, graft failure or death (whatever occurred first).

Relapse/progression incidence: Defined as the probability of having a relapse/progression within 2 years of HSCT. Patients were considered to have experienced an event when they relapsed/progressed. Death without relapse/progression and graft failures were competing risks. Patients alive with no history of relapse/progression were censored at time of last clinical examination of disease status.

Overall survival: Overall survival was defined as the probability of survival irrespective of disease status at any point in time within 2 years after HSCT.

Non-relapse mortality: NMR was defined as the probability of dying without previous occurrence of a relapse or progression within 2 years after HSCT.

Transplantation-related mortality: Transplantation-related mortality was defined as all deaths occurring due to GvHD, cardiac toxicity, pulmonary toxicity, interstitial pneumonitis, haemorrhage, hepatic sinusoidal obstruction syndrome (HSOS), skin toxicity, Epstein-Barr virus (EBV) proliferative disease, renal failure, gastrointestinal toxicity, rejection/poor graft function, central nervous system (CNS) toxicity, multiple organ failure, infections (bacterial, viral, fungal, parasitic, unknown), or other HSCT-related causes.

Engraftment: Granulocyte engraftment was documented by specifying the first of 3 consecutive days with absolute neutrophilic granulocyte count > 0.5 x 10^9/L in the peripheral blood (PB). Leukocyte engraftment was documented by specifying the first of 3 consecutive days with total leukocyte count > 1 x 10^9/L in the PB. Platelet engraftment was documented by specifying the first of 3 consecutive days with a) platelets > 20 x 10^9/L and b) platelets > 50 x 10^9/L, in the absence of platelet transfusion.

B.2.3 Summary of methodology of the relevant clinical effectiveness evidence

The combination of dose-reduced i.v. busulfan and fludarabine is one of the most frequently used and widely accepted RIC regimens, especially for patients with AML and MDS.[9,70,71] Accordingly, the selected comparator in MC-FludT.14/L Trial II was the RIC regimen i.v. busulfan (6.4 mg/kg) combined with i.v. fludarabine (150 mg/m^2), as it is considered accepted medical practice for the selected patient population. Importantly, the choice of comparator for this study was confirmed by the European Medicines Agency (EMA) during scientific advice procedures with medac. The treosulfan dose initially used in this study was 3 x 14 g/m²/day for 3 days BSA. However, following a first interim analysis and the recommendations of the independent Data Monitoring Committee, MC-FludT.14/L was substantially amended regarding the dose and schedule of the treosulfan regimen (clinical trial protocol Amendment 03). The dose of treosulfan was reduced from 3 x 14 g/m² BSA to 3 x 10 g/m² BSA, and the start of treosulfan conditioning was postponed until day-4 to make

this arm more similar to the reduced-intensity reference arm. Moreover, the follow-up of patients was extended to 2 years after transplantation and the statistical design of the comparative trial was revised. Due to the significant changes in the treosulfan dosage and regimen in Clinical Trial Protocol Amendment 03, all previously enrolled patients (N=330) were analysed separately (MC-FludT.14/L Part I) and are not included in this submission nor were they included in the statistical analyses of data collected after re-activation of the study in June 2013. Thus, the second interim analysis on 476 patients[72] and the final analysis on 570 patients (MC-FludT.14/L Trial II)[60] are considered separate to, and not a continuation of, the study conducted before implementation of protocol Amendment 03.

B.2.3.1 Trial design for MC-FludT.14/L Trial II

MC-FludT.14/L Trial II was a randomised, parallel-group, open label, multicentre, international, group-sequential phase III non-inferiority trial to evaluate efficacy and safety of treosulfan-based conditioning versus (vs) a busulfan-based RIC treatment prior to allogeneic HSCT.

In order to stop the trial as soon as the question of non-inferiority could be answered, a group-sequential approach was implemented consisting of three confirmatory interim analyses.

B.2.3.1.1 Randomisation

Patients were randomised to receive either 10 g/m² body surface area (BSA) treosulfan i.v. (test group) on Day -4, -3 and -2, or 4 x 0.8 mg/kg body weight (BW) busulfan i.v. (reference group) on Day -4 and -3 followed by allogeneic HSCT on Day 0. Randomisation was stratified by cytogenetic and/or molecular risk group for AML, or Revised International Prognostic Scoring System (IPSS-R) for MDS, and by donor type and transplantation centre.

Patients who met the enrolment criteria were centrally randomised by means of a computer-generated randomisation list to either treatment with treosulfan or busulfan. Randomisation was performed in a 1:1 ratio using a permuted block technique. The block length was unknown to the transplant centres.

To increase homogeneity between the 2 treatment arms, randomisation was stratified by cytogenetic and/or molecular risk group for AML, Revised International Prognostic Scoring System (IPSS-R) for MDS, as well as donor type:

- Risk group I: low risk and intermediate risk for AML or very low/low/intermediate IPSS-R for MDS
  Risk group II: high risk for AML and high/very high IPSS-R risk for MDS
- Donor type: MUD vs MRD.

B.2.3.1.2 Blinding

Due to different treatment schedules within the test group (treosulfan) and the reference group (busulfan) with regard to the different infusion regimens as well as the additional, anticonvulsive treatment, which was mandatory in the reference group only, blinding of the trial medication was considered unfeasible within this orphan indication of high risk patients.

In addition, transplant centres usually constitute small units with limited staff only, so that a two physician concept consisting of independent treating and rating physicians could not be implemented. The robust primary endpoint (EFS) of the trial, which is considered independent from the subjective view of the patient or the investigator, allows the conduct of the trial as an unblinded, but randomised open label trial.

Despite these considerations, sponsor and trial personnel were blinded to aggregated data evaluated during the trial. An independent contract research organisation (CRO; Clinipace, formerly Accovion) conducted the unblinded analyses needed for the Data Monitoring Committee (DMC) meetings and a statistician from this CRO presented the results at the DMC meetings. Internal statisticians and statistical programmers were unblinded after finalisation of the Statistical Analysis Plan (SAP) and data release for the final confirmatory analysis on 31-May-2017.

B.2.3.2 Eligibility criteria

B.2.3.2.1 Inclusion criteria

The patient population of this trial was selected by objective criteria, identifying patients at increased risk for standard conditioning therapies and who were,

therefore, considered to be ineligible for registered standard high-dose regimens based on busulfan or TBI.

- Patients with acute myeloid leukaemia according to the World Health Organisation (WHO) 2008,[73] (AML in complete remission at transplant, i.e. blast counts < 5% in bone marrow) or myelodysplastic syndrome according to the WHO 2008,[73] (MDS with blast counts < 20% in bone marrow during disease history) indicated for allogeneic haematopoietic progenitor cell transplantation but considered to be at increased risk for standard conditioning therapies according to the following criteria:
    - patients aged ≥ 50 years at transplant

      and/or
    - patients with a Haematopoietic Cell Transplantation-Comorbidity Index (HCT-CI) score >2 (according to Sorror et al., 2005)[7]
- Availability of a human leukocyte antigen (HLA)-identical sibling donor (MRD) or HLA-identical unrelated donor (MUD). Donor selection was based on molecular high-resolution typing (4 digits) of class II alleles of the DRB1 and DQB1 gene loci and molecular (at least) low-resolution typing (2 digits) of class I alleles (i.e., antigens) of the HLA- A, B, and C gene loci. In case no class I and class II completely identical donor (10 out of 10 gene loci) could be identified, one antigen disparity (class I) and/or one allele disparity (class II) between patient and donor was acceptable. Conversely, disparity of 2 antigens (irrespective of the involved gene loci) was not accepted. These definitions for the required degree of histocompatibility applied to the selection of related as well as unrelated donors.
- Adult patients of both gender, 18 – 70 years of age
- Karnofsky Index ≥ 60%
- Written informed consent
- Men capable of reproduction and women of childbearing potential had to be willing to consent to using a highly effective method of birth control such as condoms, implants, injectables, combined oral contraceptives, intrauterine devices, sexual abstinence or vasectomised partner while on treatment and for at least 6 months thereafter.

B.2.3.2.2 Exclusion criteria

Patients were to be excluded for any of the following reasons. Please note that the first exclusion criterion was different in France, based on the French competent authority's request:

Applied to Germany, Hungary, Italy, and Poland:

- Patients with acute promyelocytic leukaemia with t(15;17)(q22;q12) and in first complete remission (CR1)

Applied to France only: Patients with the following conditions were excluded from treatment within this trial:

- patients with acute promyelocytic leukaemia with t(15;17)(q22;q12) and in CR1
- patients with cytogenetic favourable acute myeloid leukaemia ("low risk" AML) and in CR1, who did not present unfavourable clinical or disease features like secondary or therapy-related AML or insufficient response to AML induction therapy
- MDS patients with IPSS-R "very low risk" or "low risk" at trial entry, who did not present unfavourable clinical features during disease history like refractory severe thrombocytopenia with severe bleeding complications, life-threatening infectious complications due to severe neutropenia and/or very high red blood cell transfusion requirement and related complications.

Applied to all countries:

- Patients considered contra-indicated for allogeneic HSCT due to severe concomitant illness (within 3 weeks prior to scheduled Day -6):
    - patients with severe renal impairment like patients on dialysis or prior renal transplantation or S-creatinine > 3.0 x upper limit of normal (ULN) or calculated creatinine clearance < 60 ml/min
    - patients with severe pulmonary impairment, DLCOSB (Hb-adjusted)/or forced expiratory volume 1 second (FEV_1) < 50% or severe dyspnoea at rest or requiring oxygen supply
    - patients with severe cardiac impairment diagnosed by electrocardiogram (ECG) and left ventricular ejection fraction (LVEF) < 40%

- patients with severe hepatic impairment indicated by hyperbilirubinaemia > 3 x ULN or ALT/AST > 5 x ULN.

- Active malignant involvement of the central nervous system (CNS)
- HIV-positivity, active non-controlled infectious disease under treatment (no decrease of C-reactive protein [CRP] or procalcitonin [PCT]) including active viral liver infection
- Previous allogeneic HSCT
- Pleural effusion or ascites > 1.0 L
- Pregnancy or lactation
- Known hypersensitivity to treosulfan, busulfan and/or related ingredients
- Participation in another experimental drug trial within 4 weeks prior to Day –6 of the protocol
- Non-cooperative behaviour or non-compliance
- Psychiatric diseases or conditions that might compromise the ability to give informed consent.

B.2.3.3 Settings and locations where the data were collected

The clinical trial was performed at 31 sites specialised in conducting alloHSCTs: 2 sites in France, 20 sites in Germany, 6 sites in Italy, 2 sites in Poland, and 1 site in Hungary. Patients were usually hospitalised from 1 week before start of conditioning up to successful engraftment of donor stem cells (an average for a total of 4 weeks).

The trial setting was comparable to standard NHS practice in the UK.[74]

B.2.3.4 Trial drugs and concomitant medications

B.2.3.4.1 Test drug intravenous treosulfan

Following randomisation to the test group (3 x 10 g/m² i.v. treosulfan), each patient's dose was calculated based on the patient's actual BSA (Day -4, -3 and -2) according to DuBois and DuBois.[75]

$$BSA\ (m^2) = 0.007184 \times weight\ [kg]^{0.425} \times height\ [cm]^{0.725}$$

It was accepted to round up/down the calculated BSA for one decimal digit. The administered dose of treosulfan was allowed to deviate from the calculated dose by up to 10%.

The treatment schedule for the test drug (i.v. treosulfan) is given in Table 12.

Table 12: Treatment schedule test group (intravenous treosulfan)

Visit (Day)	-6	-5	-4	-3	-2	-1	0	+1	+3	+6
Treosulfan i.v. (trial medication) (10 g/m² within 120 min)			X	X	X					
Fludarabine i.v. (30 mg/m² within 30 min)	X	X	X	X	X					
Applied to Germany, Hungary, Italy, Poland: ATG-S-Fresenius / Grafalon® i.v. (10 mg/kg in case of MUD only)			X	X	X					
Applied to France only: ATG-Thymoglobuline i.v. (2.5 mg/kg in case of MUD only)					X	X				
Allogeneic stem cell transplantation							X			
Ciclosporin i.v. daily administration (3 mg/kg/day start, 5 mg/kg/day PO)*						X	X	X	X	X
Methotrexate i.v. (mg/m²/day)								15	10	10
Calcium-Folinate i.v. (mg/m²; 6 hours after Methotrexate)								15	10	10

Source: MC-FludT.14/L Trial II[60]

* Ciclosporin dose levels adapted to the standards of the participating centre; treatment starts i.v.

ATG = anti-thymocyte globulin; i.v. = intravenous; MUD = match unrelated donor; PO = per os, oral(ly).

If both drugs were to be given on the same day, the 2-hour infusion of the total treosulfan solution was to be given prior to the infusion of fludarabine. The relation of dosing to meals was not specified.

B.2.3.4.2 Reference drug intravenous busulfan

Following randomisation to the reference group (8 x 0.8 mg/kg BW i.v. busulfan) each patient's dose was to be calculated based on the actual BW. The administered dose of busulfan was allowed to deviate from the calculated dose by up to 10%. For obese adult patients (body mass index ≥ 35 kg/m²), dosing was based on adjusted ideal BW according to the SmPC of Busilvex®.

The treatment schedule for the reference drug (i.v. busulfan) was derived from previously reported clinical data and is given in Table 13. The relation of dosing to meals was not specified.

Table 13: Treatment schedule reference group (intravenous busulfan)

Visit (Day)	-6	-5	-4	-3	-2	-1	0	+1	+3	+6
Phenytoin PO (mg) (3 x per day)*		200	100	100	100					
Busulfan i.v. (trial medication) (4 x 0.8 mg/kg/d within 120 min)			X	X						
Fludarabine i.v. (30 mg/m² within 30 min)	X	X	X	X	X					
Applied to Germany, Hungary, Italy, Poland: ATG-S-Fresenius / Grafalon® i.v. (10 mg/kg in case of MUD only)			X	X	X					
Applied to France only: ATG-Thymoglobuline i.v. (2.5 mg/kg in case of MUD only)					X	X				
Allogeneic stem cell transplantation							X			
Ciclosporin i.v. daily administration (3 mg/kg/day start, 5 mg/kg/day PO)*						X	X	X	X	X
Methotrexate i.v. (mg/m²/day)								15	10	10
Calcium-Folinate i.v. (mg/m²; 6 hours after methotrexate)								15	10	10

Source: MC-FludT.14/L Trial II[60]

* Phenytoin could be replaced by adequate benzodiazepine treatment in accordance with SmPC Busilvex®

** Ciclosporin dose levels adapted to the standards of the participating centre; treatment starts i.v.

Abbreviations: ATG = anti-thymocyte globulin; MUD = match unrelated donor; PO = per os; oral(ly); i.v. = intravenous.

B.2.3.4.3 Concomitant medications

The patients were not allowed to participate in another experimental drug trial within 4 weeks prior to Day –6 of the protocol. Due to the complexity of conditioning treatment including cytotoxic therapy, pre- and post-transplant immunosuppression and other prophylactic treatments (to prevent infections, liver, renal or CNS toxicity), relevant concomitant treatments were standardised and declared mandatory in both trial arms. In France, a different ATG preparation and regimen was used in MUD transplantation (ATG-Thymoglobuline i.v., 2.5 mg/kg on Day -2 and -1). This regimen is registered and was considered equivalent to the regimen used in the other countries. Other treatments, which were not standardised according to this section were to be conducted according to the centre-specific policy.

Relevant concomitant drug treatment given in the trial period between Day -6 and Day +28, i.e., conditioning treatments, prophylactic medication for HSOS, prophylactic medication for mucositis and growth factors (such as granulocyte colony-stimulating factor [G-CSF], granulocyte-macrophage colony-stimulating factor [GM-CSF], recombinant human kerati0nocyte growth factor [rHU-KGF]), were to be recorded on the case report form (CRF).

In addition, concomitant medication for prophylactic and/or therapeutic GvHD treatment was documented explicitly between Day -1 and Day +100 on the CRF. The restriction of the documentation of concomitant medications to these selected groups was justified by their potential impact on the primary or secondary endpoints of this trial.

Disease-specific interventions, which might have an impact on the primary trial objective (e.g., prophylactic or pre-emptive donor lymphocyte infusion (DLI), prophylactic/pre-emptive cytotoxic chemotherapy or radiotherapy after transplantation, but in the absence of relapse/disease progression), were not allowed.

B.2.3.5 Outcomes used in the economic model or specified in the scope, including primary outcome

All outcomes in the MC-FludT.14/L Trial II were pre-specified in the Statistical Analysis Plan (SAP).[76]

B.2.3.5.1 Pre-specified primary endpoint: EFS within 24 months after alloHSCT

The primary endpoint of the MC-FludT.14/L Trial II trial was event-free survival within 24 months after alloHSCT. Events were defined as relapse of disease, graft failure, or death (whatever occurred first).

B.2.3.5.2 Pre-specified secondary endpoints

1. Comparative evaluation of incidence of CTC grade III/IV mucositis between Day -6 and Day +28.

2. Comparative evaluation of overall survival (OS) and cumulative incidence of relapse (RI), non-relapse mortality (NRM) and transplantation-related mortality (TRM) within 2 years after transplantation.

3. Comparative evaluation of day +28 conditional cumulative incidence of engraftment.

4. Comparative evaluation of Day +28 and Day +100 incidence of complete donor-type chimerism.

5. Comparative evaluation of cumulative incidence of acute and chronic GvHD within 2 years after transplantation.

6. Comparative evaluation of incidence of other CTC grade III/IV adverse events between Day -6 and Day +28.

In addition, post-surveillance with respect to OS and EFS of patients who terminated the study alive at 2 years after transplantation was conducted one year after transplantation of the last randomised patient.

B.2.3.5.3 Changes in planned analyses

Two additional secondary endpoints were included in the analysis after the Blood and Marrow Transplant Clinical Trials Network recognised the potential utility of these composite endpoints in trials of allo-HCT:[77]

- GvHD-free and relapse/progression-free survival (GRFS), and
- chronic GvHD-free and relapse/progression-free survival (CRFS).

These endpoints were pre-specified in version 4.0 of the SAP prior to database lock and unblinding for the final confirmatory analysis.

Additionally, two exploratory endpoints were included in the analysis with respect to the evaluation of the Karnofsky Performance Score (KPS):

- the time to deterioration of KPS by at least 20 points, and
- the time to deterioration of KPS to less than 60 points.

These endpoints are considered as a clinically relevant description of the patient's activity and quality of life under treatment and during follow-up until 24 months after allogeneic HSCT. These endpoints were already evaluated post-hoc in version 1.0 of the clinical trial report (CTR) and were implemented in the latest SAP version 6.0 prior to database lock for the current analysis.

B.2.3.6 Summary of trial methodology

In MC-FludT.14/L Trial II, of the 570 patients randomised in this study, 551 patients qualified for the full analysis set (FAS), 537 qualified for per protocol set (PPS) and 553 qualified for safety analysis set (SAS).[60]

Table 14 summarises the trial methodology for MC-FludT.14/L Trial II.

Table 14: Comparative summary of trial methodology

Trial number (acronym)	**MC-FludT.14/L Trial II**
Location	International multi-centre trial conducted in 5 countries
Trial design	Randomised, parallel-group, open label, multicentre, group-sequential phase III non-inferiority trial
Eligibility criteria for participants	Patients with AML (in complete remission at transplant) or MDS indicated for alloHSCT but considered to be at increased risk for standard conditioning therapies
Settings and locations where the data were collected	Data were collected in centres specialised in conducting alloHSCTs. 2 sites in France, 20 sites in Germany, 6 sites in Italy, 2 sites in Poland, and 1 site in Hungary.
Trial drugs (the interventions for each group with sufficient details to allow replication, including how and when they were administered) Intervention(s) (n=[x]) and comparator(s) (n=[x]) Permitted and disallowed concomitant medication	Patients were randomised to receive either treosulfan (n=280) or busulfan (n=290) Treosulfan (test group): 10 g/m² body surface area (BSA) treosulfan i.v. on Day -4, -3 and -2 followed by alloHSCT on Day 0. Busulfan (reference group): 4 x 0.8 mg/kg body weight (BW) busulfan i.v. (reference group) on Day -4 and -3 followed by alloHSCT on Day 0. The patients were not allowed to participate in another experimental drug trial within 4 weeks prior to Day –6 of the protocol. Due to the complexity of conditioning treatment including cytotoxic therapy, pre- and post-transplant immunosuppression and other prophylactic treatments (to prevent infections, liver, renal or CNS toxicity), relevant concomitant treatments were standardised and declared mandatory in both trial arms. In France, a different ATG preparation and regimen was used in MUD transplantation. This regimen is registered and was considered equivalent to the regimen used in the other countries. Other treatments, which were not standardised were to be conducted according to the centre-specific policy. Relevant concomitant drug treatment given in the trial period between Day -6 and Day +28, i.e., conditioning treatments, prophylactic medication for HSOS, prophylactic medication for mucositis and growth factors (such as G-CSF, GM-CSF, rHU-KGF), were to be recorded on the CRF. In addition, concomitant medication for prophylactic and/or therapeutic GvHD treatment was documented explicitly between Day -1 and Day +100 on the CRF. Disease-specific interventions, which might have an impact on the primary trial objective (e.g., prophylactic or pre-emptive donor lymphocyte infusion (DLI), prophylactic/pre-emptive cytotoxic

Trial number (acronym)	MC-FludT.14/L Trial II
	chemotherapy or radiotherapy after transplantation, but in the absence of relapse/disease progression), were not allowed.
Primary outcomes (including scoring methods and timings of assessments)	Event-free survival (EFS) 2 years after transplantation; EFS was measured from time of end of HSCT (= Day 0) to time of event. Events were defined as relapse of disease, graft failure or death (whatever occurred first).
Other outcomes used in the economic model/specified in the scope	Comparative evaluation of incidence of CTC grade III/IV mucositis between Day -6 and Day +28 Comparative evaluation of overall survival (OS) and cumulative incidence of relapse (RI), non-relapse mortality (NRM) and transplantation-related mortality (TRM) within 2 years after transplantation. Comparative evaluation of cumulative incidence of acute and chronic GvHD within 2 years after transplantation. Comparative evaluation of incidence of other CTC grade III/IV adverse events between Day -6 and Day +28.
Pre-planned subgroups	Several subgroup analyses were planned including Cox proportional hazards regression model and Kaplan Meier analyses by donor type, risk group, combination of donor type and risk group, disease (AML vs MDS), age group (< 50 years vs ≥50 years), and HCT-CI Score (≤2 vs > 2)

Source: MC-FludT.14/L Trial II[60]

B.2.3.7 Baseline characteristics of trial participants

Of the 570 patients randomised in this study, the majority of patients were recruited in Germany (67.9%), with the largest number of patients included at 1 site (20.0%). At each site, patients were randomly distributed between the treosulfan and busulfan treatment groups.

Across both treatment groups, the majority of patients were male (60.8%) with a mean (SD) age of 59.6 (6.3) years. In both groups, the majority of patients were 50 years of age or older (treosulfan group, 94.0%; busulfan group, 95.8%). The age profile of the patients in the trial was representative of patients in the UK undergoing RIC.[78,79]

The mean body weight of patients was 79.4 (SD 17.7) kg in the busulfan group and 80.9 (SD 16.7) kg in the treosulfan group.

Randomisation was stratified by donor type and risk group and thus, these factors were evenly distributed to the 2 treatment groups. More patients in the trial had MUD transplantations (76.5%) than MRD (23.5%), and more patients with AML (64.0%)

than MDS (36.0%) were included in the trial. There were more patients with AML in the treosulfan group (68.6%) than the busulfan group (59.7%).There were also more patients in AML risk group II in the treosulfan group (36.1% of treosulfan patients) than the busulfan group (26.2% of busulfan patients). Overall, disease and patient characteristics in the two groups were comparable.

Table 15 summarises the characteristics of participants in the study across treatment groups.

Table 15: Baseline characteristics of participants in the studies across treatment groups in MC-FludT.14/L

Baseline characteristic - Full Analysis Set (FAS)	**Treosulfan (10 g/m²/day [d-4 to d-2])**	**Busulfan (3.2 mg/kg/day [d-4 to d-3])**
Patients analysed, N= 551	**268**	**283**
Age, mean age, years (SD)	59.3 (6.5)	59.9 (6.0)
<50 years, n (%)	16 (6.0)	12 (4.2)
≥50 years, n (%)	252 (94.0)	271 (95.8)
Male, n (%)	162 (60.4)	173 (61.1)
Female, n (%)	106 (39.6)	110 (38.9)
Indications for alloHSCT		
AML, n (%)	184 (68.6)	168 (59.4)
MDS, n (%)	84 (31.3)	115 (40.6)
Disease status – AML, n (%)		
First complete remission (CR1)	159 (86.4)	144 (85.7)
>CR1	25(13.6)	24 (14.3)
Disease status – MDS, n (%)		
Untreated	42 (50.0)	47 (40.9)
Treated	42 (50.0)	68 (59.1)
Disease duration, mean (SD)		
AML (months)	8.16 (7.54)	7.57 (7.55)
MDS (months)	14.64 (22.81)	14.01 (18.09)
HCT-CI score at baseline, n (%)		
Patients with HCT-CI score >2	156 (58.2)	167 (59.0)
Donor type, n (%)		
MRD	62 (23.1)	68 (24.0)
MUD	206 (76.9)	215 (76.0)
MUD	206 (76.9)	215 (76.0)

Source: MC-FludT.14/L Trial II[60]
Abbreviations: AML, acute myeloid leukaemia; allo, allogeneic; CR, complete remission; FAS, full analysis set; HCT-CI, haematopoietic cell transplantation co-morbidity index; HSCT, haematopoietic stem cell transplantation; MDS, myelodysplastic syndrome; MRD, matched related donor; MUD, matched unrelated donor; SD, standard deviation.

In summary, demographic characteristics were similar in the two treatment groups. Within the FAS there were more male than female patients (60.8% male, 39.2% female). In accordance with the purpose of the trial and the inclusion criteria, all patients were at increased risk for the standard HSCT conditioning regimen. The mean age of patients was 59.6 years, and 58.6% had an HCT-CI score >2. There were more patients with AML in the treosulfan group (68.6%) than the busulfan group (59.7%). However, disease status and risk group stratification for AML and MDS was comparable between the treatment groups. Stratified randomisation therefore resulted in well balanced treatment groups.

B.2.4 Statistical analysis and definition of study groups in the relevant clinical effectiveness evidence

Table 16 summarises the statistical analyses including sample size, interim analyses and stopping guidelines, statistical methods used to compare groups for primary and secondary outcomes, and the methods for additional analyses, such as subgroup analyses and adjusted analyses. Further details of the participants eligible to enter the trial can be found in Appendix D.

Table 16: Summary of statistical analyses

Trial number (acronym)	**MC-FludT.14/L Trial II**
Hypothesis objective	The primary objective of this randomised phase III trial was to demonstrate, as a minimum, non-inferiority of treosulfan as an alternative conditioning agent to busulfan with respect to EFS.
Statistical analysis	The non-inferiority margin on the hazard ratio scale was pre-specified as 1.3. If significant non-inferiority within the Per Protocol Set (PPS) could be shown, a sequential testing approach was to be applied starting with testing the non-inferiority within the Full Analysis Set (FAS). In case of statistical significance, superiority within the FAS with respect to the primary trial endpoint was to be tested based on the 'Points to Consider on Switching between Superiority and Non-inferiority (CPMP/EWP/482/99)'. For confirmatory analysis of non-inferiority of treosulfan-based conditioning a Cox proportional hazards regression model (stratified by centre and risk group) with donor type (MUD vs MRD) and treatment as factors was applied for event-free survival. These factors are exactly those factors used within the

	randomisation procedure. The analysis was based on the patients available after the second interim analysis. Beyond this confirmatory aim of the trial, exploratory data analyses were conducted. They consisted of a Cox proportional hazards regression model and Kaplan-Meier analyses by donor type, risk group, combination of donor type and risk group, disease (AML vs MDS), age group (< 50 years vs ≥ 50 years), and HCT-CI Score (≤2 vs > 2) for all patients included in the trial. To simplify descriptive comparisons, the subgroup analyses were graphically summarised by means of forest plots showing event-free survival by donor type, risk group, combination of donor type and risk group, disease (AML and MDS), age group (< 50 years vs ≥ 50 years), HCT-CI Score (≤2 vs >2), remission status in AML (CR1 vs > CR1), disease status at trial entry in MDS (untreated vs treated), risk group within AML patients, and risk group within MDS patients. For each subgroup they include the associated sample size, the number of events, the 24-month Kaplan-Meier estimates, the hazard ratio (HR) showing the risk of events with treosulfan compared to busulfan and the associated CI derived by means of Cox proportional hazards model with treatment as only factor. Sensitivity analyses were also performed to evaluate the robustness of the primary analysis. For the secondary endpoints, data for the FAS and PPS were analysed. For the endpoints overall survival, transplantation-related mortality, GvHD-free and relapse/progression-free survival, and chronic GvHD-free and relapse/progression-free survival and the exploratory endpoints time to deterioration of Karnofsky Performance Score (KPS) by at least 20 points and deterioration of KPS to less than 60 points, Kaplan-Meier estimates were calculated. A Cox proportional hazards regression model stratified by centre and risk group with donor type (MUD vs MRD) and treatment as factors was fitted. The adjusted estimate of treatment effect from the Cox proportional hazards model was expressed as a hazard ratio (treosulfan vs busulfan), together with the associated 95% CIs. In addition, the two-sided p-values based on Wald-test was conducted. For the endpoints relapse/progression, and non-relapse mortality, the probability over time was estimated by cumulative incidence rates. The test of Gray was applied to compare treatment arms. For engraftment, the conditional cumulative incidence was estimated using conditional probability functions. The two-sided Pepe-Mori test was used to compare treatment arms. Potential effects of covariates on the secondary endpoints were studied through subgroup analyses. The same subgroups and analysis techniques described above for event-free survival were considered. However, the non-inferiority testing and superiority testing was omitted because the trial was designed to test non-inferiority with respect to event-free survival only. The trial was planned as a group-sequential trial with 3 interim analyses. The first formal interim analysis was planned to be performed after 45 events or 220 patients to allow for a broad review of the benefits and risks of the dose reduction and change of the treatment regimen implemented with amendment 3 of the trial. Stopping early due to proof of efficacy or futility was unlikely within this interim analysis due to the low information fraction within this analysis. Further interim analyses were planned after 137 and 239 events occurred, or after 460 and 700 patients were randomised. The final analysis was planned after 481 events or inclusion of 930 patients at the latest. When reviewing the results of the second interim analysis, the DMC recommended to stop recruitment for this trial since the primary objective – the proof of non-inferiority of treosulfan compared to busulfan – had already been achieved. The Clinical

	Study Report (CSR) describes the results of the final analysis of all 570 patients enrolled in the trial and of these, 551 patients qualified for the FAS.
Sample size, power calculation	The sample size required within the scope of amendment 03 was calculated based on the hypothesis system described in the CSR (Section 9.7.1.1) applying an experiment-wise one-sided type-I-error significance α = 2.5%. Only those patients to be enrolled after implementation of the amendment 03 were subjected to confirmatory analysis, i.e. the 330 patients already recruited in MC-FludT.14/L Trial II prior to trial re-activation with amendment 03 were excluded. The rationale for this is twofold: All patients previously randomised to the treatment regimen with 14 g/m^2 treosulfan do not provide information for safety and efficacy of the newly developed regimen with 10 g/m^2. In addition, all patients previously randomised to busulfan may not be representative for future randomised patients due to potential selection- and performance-bias. Exclusion of any kind of bias is even more important in non-inferiority design settings. Sample size estimation assumed under the alternative hypothesis that treosulfan-based conditioning is equally effective to the comparator (i.e.: HR = 1). The power of the trial was 80%, so that the sponsor´s risk of erroneously overlooking the non-inferiority was 20%. Interim analyses for futility looks were incorporated in the revised protocol allowing for premature stop of the trial if it was unlikely to achieve the ultimate goal of the trial. In addition, the trial was to be stopped early if non-inferiority of treosulfan-based conditioning was clearly established. This resulted in a group-sequential approach with at most 3 confirmatory interim analyses and one final analysis, each with different stopping criteria (boundaries) for futility and efficacy. The resulting inflation of the overall Type I and Type II error probabilities was taken into account. The most conservative approach of interim efficacy monitoring by means of an O'Brien-Fleming type stopping boundary was applied. Based on these general conditions above, a commitment to at most 481 events within the FAS was given in the trial protocol. Since the power of any time-to-event trial is determined by the number of events rather than the number of patients, a range of sample sizes met the objectives of this trial. Naturally, the more patients being followed, the sooner the desired number of events is observed. Assuming, a 12-month EFS-rate of 68.5% with busulfan-based conditioning RIC (based on the results of first confirmatory interim analysis of MC-FludT.14/L Trial II prior to amendment 03) and anticipating (based on the accrual experience prior to amendment 03 and the already existent infrastructure of the trial) recruitment of 10 patients per month within the first 6 months, 15 patients per month thereafter until 24 months after re-start of the trial and 25 patients per month thereafter, the required number of events was expected to be reached with at most 930 patients. The maximum expected trial duration (accrual plus follow-up) to reach the required number of events was about 64 months after randomisation of the first patient with amendment 03 in force. The expected duration was 40 months under the null hypothesis and 58 months under the alternative hypothesis of non-inferiority. Assuming that roughly 3% of patients have to be excluded from FAS, at most 960 patients were to be enrolled in this trial.
Interim analysis	In order to stop the trial as soon as the question of non-inferiority could be answered, a group-sequential approach was implemented consisting of 3 confirmatory interim analyses. Interim analyses were conducted to allow for early stopping of the trial for significant non-inferiority as well as futility.

	In particular, the first interim analysis was performed with 220 randomised patients qualifying for FAS to investigate the effect of dose reduction on duration of neutropenia and TRM until Day +100. The second interim analysis was scheduled with 137 events or latest after randomisation of 460 patients qualifying for FAS. The third interim analysis was scheduled with 239 events or 700 patients in FAS. After the second interim analysis, the DMC recommended stopping recruitment into the trial since the primary objective of the trial, the proof of non-inferiority of treosulfan compared to busulfan had been achieved. Medac followed this recommendation and stopped recruitment after a total of 570 patients had been included. A confirmatory analysis based on 476 patients included in the second interim analysis was completed and presented in a previous clinical trial report. The final analysis of all 570 patients included in the trial, was performed after all patients had been followed-up for at least one year and when the post-surveillance documentation had been performed for patients who finished the trial alive after 2 years.
Data management, patient withdrawals	Patients had to be permanently removed from the trial if they withdrew their consent. Only 5 of 570 patients withdrew their consent. Four of these patients are included in the FAS. The other patient withdrew his consent prior to the start of treatment. If there was a permanent removal of a patient from the trial, all efforts were to be made to perform all assessments scheduled for the end-of-trial visit. The reason for withdrawal was documented on the CRF.

Source: MC-FludT.14/L Trial II[60]

B.2.4.1 Analysis Populations

B.2.4.1.1 Safety Analysis Set

The Safety Analysis Set (Safety Set) included all randomised patients who were treated at least one time with trial medication. All patients were analysed within their group of actual treatment.

B.2.4.1.2 Full Analysis Set

The FAS included all randomised patients of the Safety Set with at least one efficacy parameter documented after baseline. The patients within the FAS were analysed in their initial group of randomisation (intention to treat principle).

B.2.4.1.3 Per Protocol Set (PPS)

The PPS comprised all patients of the FAS, if the following criteria were additionally met:

- All of the inclusion criteria, none of the exclusion criteria were fulfilled.

- Correct allocation to treatment group.
- Compliance with respect to the administration of the trial medication. Patients with a deviation of at most plus/minus 20% between the amount of actually applied and the amount of protocol required trial medication were considered as compliant.
- Administration of short course methotrexate until Day +6 unless medical reasons for a deviation had been documented.
- Administration of ATG in case of MUD unless medical reasons for a deviation had been documented.
- Lack of any concomitant prophylactic/adjuvant DLI or cytotoxic therapy/radiotherapy after transplantation but in the absence of relapse/disease progression.

B.2.5 Quality assessment of the relevant clinical effectiveness evidence

B.2.5.1 MC-FludT.14/L Trial II

Table 17 summarises the assessment of the risk of bias for the RCT, MC-FludT.14/L Trial II. Overall, the risk of bias was considered to be low given that randomisation was robust and blinding was not possible for patients or treatment providers. There were no unexpected imbalances in withdrawals during the course of the study. Furthermore, the planned outcome measures were analysed and reported and most efficacy analyses used an intention to treat approach.

Table 17: Quality assessment results for the RCT - MC-FludT.14/L Trial II

Study question	How is the question addressed in the study?	Risk of bias (High, Low, Unclear)
Was randomisation carried out appropriately?	Randomisation was performed using a robust, validated block approach. The block length was unknown to transplant centres. To increase homogeneity between the 2 treatment arms, randomisation was stratified by cytogenetic and/or molecular risk group for AML, IPSS-R for MDS, as well as donor type.	Low
Was the concealment of treatment allocation adequate?	Due to different treatment schedules with regard to the different infusion regimens as	Low

Study question	How is the question addressed in the study?	Risk of bias (High, Low, Unclear)
	well as the additional, anticonvulsive treatment, which was mandatory in the reference group only, blinding of the trial medication was considered unfeasible within this orphan indication of high risk patients. In addition, transplant centres usually constitute small units with limited staff only, so that a two physician concept consisting of independent treating and rating physicians could not be implemented. The robust primary endpoint (EFS) of the trial, which is considered independent from the subjective view of the patient or the investigator, allows the conduct of the trial as an unblinded, but randomised open label trial. The sponsor and trial personnel were blinded to aggregated data evaluated during the trial. An independent CRO conducted the unblinded analyses needed for the DMC meetings and a statistician from this CRO presented the results at the DMC meetings. Internal statisticians and statistical programmers were unblended after finalisation of the Statistical Analysis Plan (SAP) and data release for the final confirmatory analysis on 31-May-2017.	
Were the groups similar at the outset of the study in terms of prognostic factors, for example, severity of disease?	Demographic characteristics were similar in the 2 treatment groups. Within the FAS there were more male than female patients (60.8% male, 39.2% female). There were more patients with AML in the treosulfan group (68.6%) than the busulfan group (59.7%). However, disease status and risk group stratification for AML and MDS was comparable between the treatment groups. Stratified randomisation	Low

Study question	How is the question addressed in the study?	Risk of bias (High, Low, Unclear)
	therefore resulted in well balanced treatment groups.	
Were the care providers, participants and outcome assessors blind to treatment allocation? If any of these people were not blinded, what might be the likely impact on the risk of bias (for each outcome)?	Participants and healthcare providers could not be blinded as previously explained. However, sponsor and trial personnel were blinded to aggregated data evaluated during the trial. An independent CRO conducted the unblinded analyses needed for the DMC meetings and a statistician from this CRO presented the results at the DMC meetings.	Low
Were there any unexpected imbalances in drop-outs between groups? If so, were they explained or adjusted for?	No	Low
Is there any evidence to suggest that the authors measured more outcomes than they reported?	No	Low
Did the analysis include an intention-to-treat analysis? If so, was this appropriate and were appropriate methods used to account for missing data?	Yes, the patients within the FAS were analysed in their initial group of randomisation (intention to treat principle).	Low
Did the authors of the study publication declare any conflicts of interest?	Study is unpublished.	

Source: MC-FludT.14/L Trial II[60]
CSR, clinical study report; FAS, full analysis set; RCT, randomised controlled trial.

B.2.5.2 Comparison of MC-FludT.14/L Trial II and routine clinical practice in England

There is limited direct evidence supporting the current practice in selecting patients with a wide range of malignancies (e.g. AML, ALL and MDS), for RIC versus MAC regimens.[6] According to the EBMT, current strategies should focus on the concept that patients with a high risk for TRM and a low disease risk should receive a different conditioning regimen from patients with a low risk for TRM and high risk disease.[21]

The worldwide data (2007-20117) from the Center for International Blood and Marrow Transplant Research (CIBMTR) indicate that total body irradiation (TBI), combinations of busulfan with cyclophosphamide or fludarabine, and fludarabine combined with melphalan are among the most commonly used conditioning regimens for patients with AML or MDS undergoing myeloablative or RIC alloHSCT (Figure 3).[54]

The MC-FludT.14/L study closely reflects routine clinical practice in the UK which is illustrated in Killick et al 2014,[80] The National Cancer Research Institute's protocols for AML (NCRI AML15 2004;[81] NCRI AML16 2006;[82] NCRI AML19)[83] and the 2016 Yorkshire and Humberside Clinical Networks guidelines for AML.[84]

B.2.6 Clinical effectiveness results of the relevant trials

Clinical effectiveness from MC-FludT.14/L Trial II

- MC-FludT.14/L Trial II is the pivotal phase III RCT assessing the safety and efficacy of treosulfan in combination with fludarabine for conditioning therapy as part of alloHSCT in adults with AML or MDS. With 570 randomised patients it was the largest ever prospective RCT comparing two conditioning regimens.
- The trial demonstrated that the treosulfan/fludarabine regimen was non-inferior to busulfan/fludarabine with regard to the primary endpoint (EFS at 24 months after alloHSCT), and that the treatment difference for EFS was statistically significantly in favour of treosulfan:[60]
    - The EFS at 24 months after alloHSCT was 65.7% (95% CI; 59.5, 71.2) for the treosulfan treatment group, and 51.2% (95% CI; 45.0, 57.0) for the busulfan treatment group, with a hazard ratio (HR) of 0.64 in favour of treosulfan (p=0.0000001)
    - The treatment difference for EFS between regimens was statistically significantly in favour of treosulfan (p=0.0005787), indicating a clinically relevant long-term advantage of treosulfan
- The secondary endpoints assessed were: engraftment, complete donor-type chimerism, OS, relapse/progression, NRM, TRM, GRFS, CRFS, acute and chronic GvHD, AEs, and Grade III/IV mucositis.
- The statistically significant advantage of the treosulfan/fludarabine regimen compared with the busulfan/fludarabine regimen was demonstrated for the majority of the secondary endpoints:[60]
    - The incidence of complete donor-type chimerism was statistically significantly higher with the treosulfan-based regimen both at day +28 (93.2% vs 83.3%; p=0.0159) and at day +100 (86.1% vs 80.2%; p=0.0381).

- The Kaplan-Meier estimate of OS at 24 months was significantly superior for patients treated with the treosulfan-based regimen than those treated with the busulfan-based regimen (72.7% vs 60.2%; p=0.0037).
- The cumulative incidence of NRM at 24 months after alloHSCT was statistically significantly lower in the treosulfan treatment group compared with the busulfan treatment group (12.0% vs 20.4%; p=0.0343).
- A statistically significant reduction in cumulative incidence of TRM at 24 months after alloHSCT was observed for the treosulfan-based regimen vs the busulfan-based regimen (12.8% vs 24.1%; p=0.0043).
- The rates of GRFS and CRFS were both statistically significantly higher in patients treated with the treosulfan-based regimen vs those treated with the busulfan-based regimen (GRFS, 50.3% vs 37.1%; p=0.0087; CRFS, 51.4% vs 37.2%; p=0.0030).
- Engraftment rates at day +28 were similar between the treosulfan-based regimen and the busulfan-based regimen (96.2% vs 96.8%; p=0.4235).
- Notably, the cumulative incidence of relapse/progression at 2 years after transplantation was comparable between the two conditioning regimens (treosulfan, 22.0%; busulfan, 25.2%; p=0.2631).

- The robustness of these results was confirmed by exploratory subgroup analyses (i.e. Kaplan Meier estimates of EFS by donor type, risk and age group, disease, HCT-CI score, and combinations thereof) and pre-planned sensitivity analyses (i.e. Cox regression models with different prognostic subgroups as factors or strata).[60]
- In the post-surveillance evaluation, follow-up data for EFS, OS, relapse/progression, and NRM showed a continued clinically relevant long-term advantage of treosulfan compared with busulfan.[60]

B.2.6.1 MC-FludT.14/L Trial II efficacy results

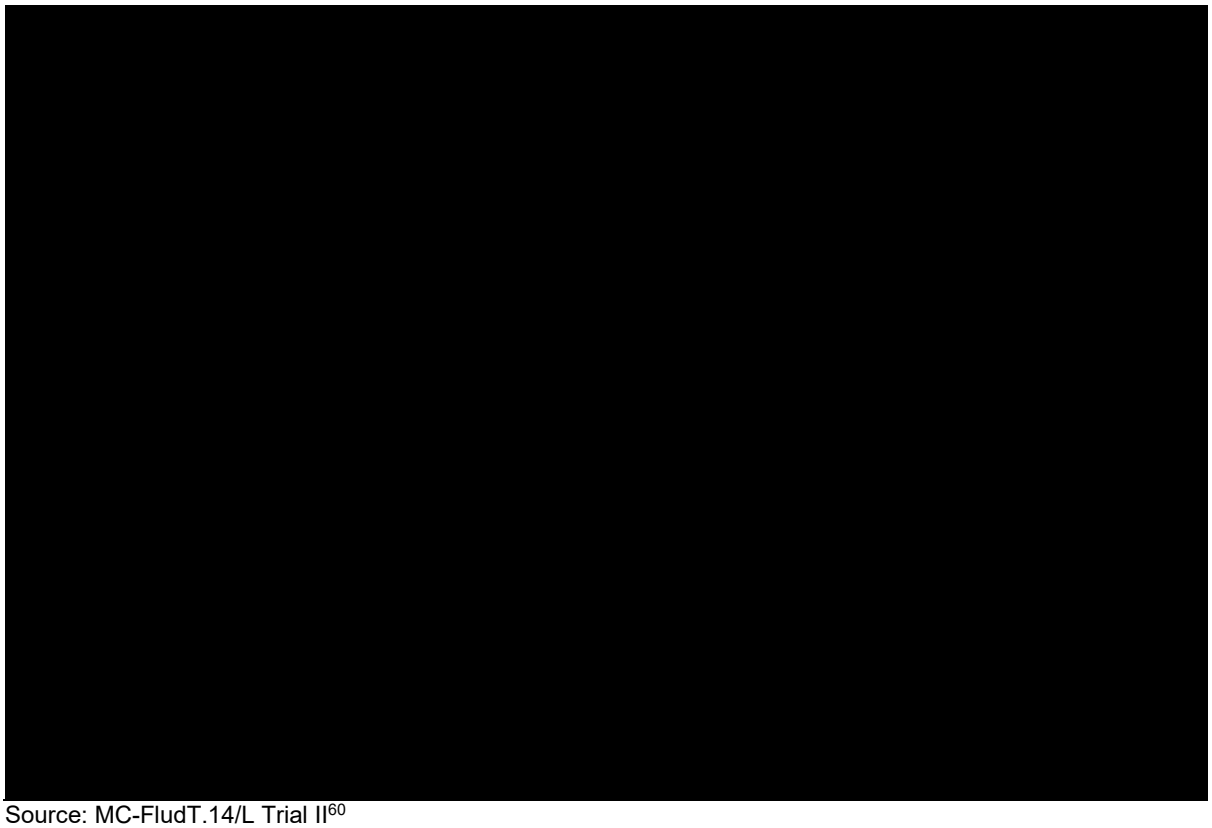
B.2.6.1.1 Primary endpoint: EFS within 24 months after alloHSCT

Event-free survival within 24 months after alloHSCT was defined as the primary endpoint of MC-FludT.14/L Trial II. Events were defined as relapse of disease, graft failure, or death (whatever occurred first).

As the non-inferiority of treosulfan to busulfan with respect to EFS was demonstrated both in the PPS and FAS populations, data throughout this dossier are presented by FAS population.

This analysis showed that the treosulfan-based regimen was non-inferior to the busulfan-based regimen with regard to the primary endpoint EFS, and that the treatment difference for EFS was statistically significantly in favour of treosulfan. In total, 36.2% of patients in the treosulfan treatment group and 48.4% of patients in the busulfan treatment group experienced an event (Table 18). The Kaplan-Meier estimate of EFS 24 months after alloHSCT was 65.7% (95% CI; 59.5, 71.2) for the treosulfan treatment group, and 51.2% (95% CI; 45.0, 57.0) for the busulfan treatment group (Figure 4). The Kaplan-Meier estimate of EFS at 36 months after alloHSCT was [redacted]% (95% CI; [redacted], [redacted]) for the treosulfan treatment group, and [redacted]% (95% CI; [redacted], [redacted]) for the busulfan treatment group. The HR was [redacted] (95% CI; [redacted], [redacted]) in favour of the treosulfan treatment group. The test for non-inferiority of treosulfan compared with busulfan resulted in a one-sided p-value of [redacted] (adjusted for donor type as a factor and risk group and centre as strata using a Cox regression model). The p-value for superiority testing of treosulfan compared with busulfan was [redacted], and the p-value for testing difference was [redacted] indicating a statistically significant and clinically relevant long-term advantage of treosulfan.

Figure 4: Kaplan Meier estimates of EFS (FAS) – MC-FludT.14/L Trial II



Source: MC-FludT.14/L Trial II[60]

Abbreviations: CI, confidence interval; FAS, full analysis set
a Adjusted for donor type as factor, and risk group and centre as strata using Cox regression model.
b For testing non-inferiority of treosulfan compared with busulfan.
c For testing superiority of treosulfan compared with busulfan.

Table 18: Primary endpoint: EFS (FAS) – MC-FludT.14/L Trial II

	Busulfan (N=283)	**Treosulfan (N=268)**
Median follow-up[a] [months] (range of those surviving)	29.4 (3.0, 54.3)	29.7 (3.0, 52.1)
Patients with event	137 (48.4%)	97 (36.2%)
Death[b]	56 (19.8%)	35 (13.1%)
Relapse/Progression[b]	72 (25.4%)	61 (22.8%)
Primary Graft Failure[b]	1 (0.4%)	1 (0.4%)
Secondary Graft Failure[b]	8 (2.8%)	0 (0.0%)
Patients without event	146 (51.6%)	171 (63.8%)
Event-free survival at 12 months[c] [%] (95% CI)	60.8 (54.9, 66.3)	70.0 (64.1, 75.1)
Event-free survival at 24 months[c] [%] (95% CI)	51.2 (45.0, 57.0)	65.7 (59.5, 71.2)
Event-free survival at 36 months[c] [%] (95% CI)	[redacted] ([redacted],[redacted])	[redacted] ([redacted],[redacted])
Hazard Ratio (treosulfan/busulfan)d (95% CI)	0.64 (0.49, 0.84)	
p-valued[e] for testing non-inferiority of treosulfan compared to busulfan	0.0000001	
p-valued for testing superiority of treosulfan compared to busulfan	0.0005787	
p-valued for testing difference	0.0011574	

Source: MC-FludT.14/L Trial II[60]
CI = confidence interval; N = number of patients.
a Based on reverse Kaplan-Meier estimates for overall survival
b Only if this event occurred first
c Based on Kaplan-Meier estimates
d Adjusted for donor type as factor, and risk group and centre as strata using Cox regression model
e The non-inferiority margin for the hazard ratio is 1.3.

The overall benefit of treosulfan in terms of EFS was consistently observed in nearly all the exploratory subgroup analyses (i.e. by donor type, risk group [I vs II], combination of donor type and risk group, disease [AML vs MDS], age group [<50 years vs ≥50 years], and HCT-CI Score [≤2 vs >2]). The forest plot for EFS by prognostic factors with 24 months event rates for the FAS is presented in Figure 5.

Figure 5: Forest plot for EFS by prognostic factors with 24 month event rates (FAS) – MC-FludT.14/L Trial II

Source: MC-FludT.14/L Trial II[60]
Abbreviations: AML, acute myeloid leukaemia; CI, confidence interval; CR, complete remission; HCT-CT, hematopoietic cell transplantation-comorbidity index; HR, hazard ratio; MDS, myelodysplastic syndrome; MRD, matched related donor; MUD, matched unrelated donor; N, number of patients; n, total number of events; RG, risk group.
* Adjusted for donor type as factor, and risk group and centre as strata using Cox regression model.

The results of the subgroup analyses were consistent with the analysis of the total population. Other than matched related donor (MRD) risk group II, HRs for each subgroup were in favour of treosulfan (HR<1.0). As the confidence intervals (CIs) for each prognostic factor or combination were widely overlapping, no differential effects could be identified.

Cox regression models with different prognostic subgroups as factors or strata were performed as pre-planned sensitivity analyses. Results from these sensitivity analyses confirmed the robustness of the primary analysis.

Results from the primary endpoint analysis based on 570 patients are consistent with those from the interim analysis based on 476 patients (FludT.14/L Trial II CSR 476 patients).[72]

In summary, EFS following treosulfan conditioning was statistically non-inferior to busulfan (adjusted p=0.0000164, Cox regression model as above) with Kaplan-Meier EFS rates at 24 months after HSCT of 50.4% (95% CI: 42.8%, 57.5%) for the busulfan treatment group, and 64.0% (95% CI: 56.0%, 70.9%) for the treosulfan treatment group. The HR was 0.65 (95% CI: 0.47, 0.90) in favour of the treosulfan treatment group. The p-value for the test of superiority was 0.0051268 (adjusted for donor type as a factor and risk group and centre as strata using a Cox regression model). Although this p-value was very low, and the difference between treatment groups was regarded as clinically relevant, the p-value in the final confirmatory analysis did not meet the significance level resulting from an O'Brien-Fleming type of group-sequential efficacy stopping boundary, which was 0.000149. Sensitivity analyses with various combinations of factors and strata included in the statistical model confirmed the robustness of the primary analysis. The overall benefit of treosulfan with regard to EFS was also observed in the exploratory subgroup analyses (i.e., by donor type, risk group, combination of donor type and risk group, disease (AML vs MDS), age group (< 50 years vs ≥ 50 years), and HCT CI Score (≤ 2 vs > 2), where the HRs were in favour of treosulfan for almost all subgroups.

The data from this final analysis of all 570 patients enrolled in the trial are consistent with those data previously reported in the confirmatory analysis.[59] The mean follow-up time was 29.4 months in the busulfan treatment group, and 29.7 months in the treosulfan treatment group, considerably longer than the 17.4 months (busulfan) and 15.4 months (treosulfan) in the confirmatory analysis. For the primary endpoint, EFS at 24 months in the PPS was 51.1% (95% CI: 44.8, 57.0) for the busulfan treatment group, and 65.3% (95% CI: 59.0, 70.9) for the treosulfan treatment group. The HR was 0.64 (95% CI: 0.48, 0.84) in favour of the treosulfan treatment group.

This final analysis included a post-surveillance evaluation for patients who completed the Month 24 visit alive. The post-surveillance evaluation was completed for 107 patients in the busulfan treatment group, and 125 patients in the treosulfan treatment group. The post-surveillance evaluation collected data for the primary endpoint EFS, and secondary endpoints relating to the primary endpoint (OS, relapse/progression, graft failure, and NRM). For each of these endpoints, in addition to the standard analysis for all endpoints, event rates were calculated for 36 months after HSCT.

Event-free survival at 36 months was [redacted]% (95% CI: [redacted], [redacted]) for the busulfan treatment group, and [redacted]% (95% CI: [redacted], [redacted]) for the treosulfan treatment group. Non-inferiority was demonstrated in the confirmatory analysis based on [redacted] patients. Therefore, no confirmatory test for non-inferiority was conducted in the present analysis. However, statistical testing in this final data analysis gave an adjusted p-value of [redacted] (FAS, adjusted for donor type as a factor and risk group and centre as strata using a Cox regression model). The p-value for exploratory testing of superiority of treosulfan compared to busulfan was [redacted] (FAS) in favour of treosulfan.

Direct comparison of MC-FludT.14/L Trial II survival data with published transplant outcome data is difficult, because primary and secondary endpoints like EFS and OS are affected by a number of trial-specific patient-, donor-, and disease-characteristics. These characteristics are in general different between MC-FludT.14/L Trial II and the low number of other published prospectively randomised phase III trials in AML and MDS patients.

Given the low number of other relevant published phase III trials, medac commissioned a matched pair analysis of the EBMT Registry data with the objectives of comparing the registry data for OS, cumulative incidence of relapse (RI) and NRM in patients undergoing alloHSCT with fludarabine/melphalan conditioning or busulfan/cyclophosphamide conditioning compared to treatment with treosulfan-based conditioning (from the MC-FludT.14/L Trial II results).[69]

This analysis supported the MC-FludT.14/L Trial II and found that for MDS patients treatment utilising treosulfan-based conditioning provided an improvement in OS (HR

[redacted] (95% CI [redacted] -[redacted]), p=[redacted] vs BuCy) and a numerically lower relapse incidence (RI) as well as NRM. The differences were clinically relevant with over [redacted]% difference for OS, up to [redacted]% difference for NRM, [redacted]% difference for RI.[69]

Similarly for AML patients, OS is significantly improved with treosulfan-based conditioning [HR [redacted] (95% CI [redacted] - [redacted]), p<[redacted] vs FluMel; HR [redacted] ([redacted] - [redacted]), p=[redacted] vs BuCy], while NRM is significantly reduced [HR [redacted] ([redacted] - [redacted]), p=[redacted] vs FluMel; HR [redacted] ([redacted] - [redacted]), p= [redacted] vs BuCy] with the use of fludarabine/treosulfan based conditioning compared to FluMel and BuCy.[69]

B.2.6.1.2 Interim analysis of primary endpoint (n=476)

An interim analysis of the primary endpoint (EFS within 24 months after alloHSCT) was conducted on 476 patients, as part of MC-FludT.14/L Trial II, prior to the final analysis based on 570 patients. Results from the interim analysis of the primary endpoints are summarised below.

In the interim analysis, EFS following treosulfan conditioning was statistically non-inferior to busulfan (adjusted one-sided p-value=0.0000164), with Kaplan Meier EFS estimates at 24 months after alloHSCT of 64.0% (95% CI; 56.0%, 70.9%) for the treosulfan treatment group, and 50.4% (95% CI; 42.8%, 57.5%) for the busulfan treatment group. The HR was 0.65 (95% CI; 0.47, 0.90) in favour of the treosulfan treatment group. The statistical non-inferiority of treosulfan to busulfan was demonstrated as the one-sided p-value of 0.0000164 (Cox regression model as above) was well below the one-sided significance level of 0.000149. The clinically relevant advantage of treosulfan was demonstrated by the low adjusted p-value of 0.0051268 (Cox regression model as above) for testing superiority; however, the p-value was slightly higher than the significance level specified for interim testing (p=0.000149).[60]

B.2.6.1.3 Secondary endpoints

For MC-FludT.14/L Trial II, the secondary endpoints were:

- Comparative evaluation of incidence of CTC grade III/IV mucositis between Day -6 and Day +28 (see Section B.2.10).
- Comparative evaluation of overall survival (OS) and cumulative incidence of relapse (RI), non-relapse mortality (NRM) and transplantation-related mortality (TRM) within 2 years after transplantation.

- Comparative evaluation of day +28 conditional cumulative incidence of engraftment.
- Comparative evaluation of Day +28 and Day +100 incidence of complete donor-type chimerism.
- Comparative evaluation of cumulative incidence of acute and chronic GvHD within 2 years after transplantation (see Section B.2.10).
- Comparative evaluation of incidence of other CTC grade III/IV adverse events between Day -6 and Day +28 (see Section B.2.10).

In addition, post-surveillance with respect to OS and EFS of patients who terminated the study alive at 2 years after transplantation was conducted one year after transplantation of the last randomised patient.

Results from the secondary endpoints analysis are summarised below. Those secondary endpoints considered as safety endpoints can be found in Section B.2.10).

Comparative evaluation of overall survival (OS)

In the FAS, 171 patients (60.4%) in the busulfan treatment group, and 187 patients (69.8%) in the treosulfan treatment group were alive at the time of the post-surveillance evaluation (Table 19). The Kaplan-Meier estimate for OS 24 months after HSCT was 60.2% (95% CI: 54.0, 65.8) for the busulfan treatment group, and 72.7% (95% CI: 66.8, 77.8) for the treosulfan treatment group.[60]

The Kaplan-Meier estimate of OS 36 months after HSCT was [redacted]% (95% CI: [redacted], [redacted]) for the busulfan treatment group, and [redacted]% (95% CI: [redacted], [redacted]) for the treosulfan treatment group. The difference between the treatment groups was statistically significant (p=0.0037, adjusted for donor type as a factor, and risk group and centre as strata using a Cox regression model). The HR was 0.64 (95% CI: 0.48, 0.87) in favour of treosulfan.[60]

Table 19: Summary results of overall survival (FAS) - MC-FludT.14/L Trial II

	Busulfan (N=283)	**Treosulfan (N=268)**
Patients with event	112 (39.6%)	81 (30.2%)
Patients without event	171 (60.4%)	187 (69.8%)
Overall survival at 12 months[a] [%] (95% CI)	71.8 (66.1, 76.7)	77.8 (72.3, 82.3)
Overall survival at 24 months[a] [%] (95% CI)	60.2 (54.0, 65.8)	72.7 (66.8, 77.8)

	Busulfan (N=283)	Treosulfan (N=268)
Overall survival at 36 months[a] [%] (95% CI)	[redacted] ([redacted],[redacted])	[redacted] ([redacted],[redacted])
Hazard Ratio (treosulfan/busulfan)[b] (95% CI)	0.64 (0.48, 0.87)	
Adjusted p-value[b]	0.0037	
p-value[c]	0.0189	

Source: MC-FludT.14/L Trial II[60]

[a] based on Kaplan-Meier estimates

[b] adjusted for donor type as factor, and risk group and centre as strata using Cox regression model

[c] Log-rank test

In the disease subgroups, OS for AML patients (FAS) was [redacted]% in the busulfan treatment group ([redacted] patients), and [redacted]% in the treosulfan treatment group ([redacted] patients). For MDS patients, OS was [redacted]% in the busulfan treatment group ([redacted] patients), and [redacted]% in the treosulfan treatment group ([redacted] patients). In the AML remission status subgroup, OS for patients in CR1 was [redacted]% in the busulfan treatment group ([redacted] patients), and [redacted]% for the treosulfan treatment group ([redacted] patients). The results of the subgroup analyses were consistent with the analysis of the total population. The HRs for MRD, MRD risk group II, and MDS risk group I were in favour of busulfan (HR > 1.0). All other HRs for each subgroup were in favour of treosulfan (HR < 1.0). As the CIs for each prognostic factor or combination were widely overlapping, no differential effects could be identified.

In summary, the overall survival at 24 months after alloHSCT was statistically significantly higher in the treosulfan treatment group compared with the busulfan treatment group (72.7% vs 60.2%; HR 0.64, 95% CI; 0.48, 0.87; adjusted p-value=0.0037). At the time of the post-surveillance evaluation, a higher proportion of patients in the treosulfan treatment group were alive compared with patients in the busulfan treatment group (69.8% vs 60.4%). The Kaplan-Meier estimate of OS at 36 months after alloHSCT was [redacted]% (95% CI; [redacted], [redacted]) for the treosulfan treatment group, and [redacted]% (95% CI; [redacted], [redacted]) for the busulfan treatment group.

Cumulative incidence of relapse (RI)

A summary of relapse/progression, is presented in Table 20 for the FAS. In the FAS, 72 patients (25.4%) in the busulfan treatment group and 61 patients (22.8%) in the treosulfan treatment group experienced relapse/progression. The cumulative incidence of relapse/progression 24 months after HSCT was 25.2% (95% CI: 20.0%,

30.3%) in the busulfan treatment group, and 22.0% (95% CI: 16.9%, 27.1%) in the treosulfan treatment group. The cumulative incidence 36 months after HSCT was [redacted]% (95% CI: [redacted], [redacted]) for the busulfan treatment group, and [redacted]% (95% CI: [redacted], [redacted]) for the treosulfan treatment group. The difference between the treatments groups was not statistically significant (p=0.2631, adjusted for donor-type as factor, and risk group as stratum using the Fine and Gray model). The HR was 0.82 (95% CI: 0.59, 1.16) in favour of treosulfan.

Table 20: Summary table of relapse/progression (FAS) - MC-FludT.14/L Trial

	Busulfan (N=283)	**Treosulfan (N=268)**
Patients with event	72 (25.4%)	61 (22.8%)
Patients without event (censored) or with competing event	211 (74.6%)	207 (77.2%)
Censored	146 (51.6%)	171 (63.8%)
Death[a]	56 (19.8%)	35 (13.1%)
Primary Graft Failure[a]	1 (0.4%)	1 (0.4%)
Secondary Graft Failure[a]	8 (2.8%)	0 (0.0%)
Cumulative incidence at 12 months [%] (95% CI)	21.7 (16.9, 26.5)	19.1 (14.4, 23.8)
Cumulative incidence at 24 months [%] (95% CI)	25.2 (20.0, 30.3)	22.0 (16.9, 27.1)
Cumulative incidence at 36 months [%] (95% CI)	[redacted] ([redacted],[redacted])	[redacted] ([redacted],[redacted])
Hazard Ratio (treosulfan/busulfan)[b] (95% CI)	0.82 (0.59, 1.16)	
Adjusted p-value[b]	0.2631	
p-value[c]	0.4271	

Source: MC-FludT.14/L Trial II[60]

[a] Only if this event occurred first

[b] Adjusted for donor type as factor and risk group as stratum using Fine and Gray model

[c] Based on test of Gray

Non-relapse mortality (NRM)

A summary of NRM data is presented in Table 21 for the FAS. In the FAS, 56 patients (19.8%) in the busulfan treatment group and 35 patients (13.1%) in the treosulfan treatment group died without relapse/progression. The cumulative incidence of NRM 24 months after HSCT was 20.4% (95% CI: 15.5%, 25.2%) in the busulfan treatment group, and 12.0% (95% CI: 8.0, 15.9) in the treosulfan treatment group. The cumulative incidence 36 months after HSCT was [redacted] (95% CI: [redacted], [redacted]) in the busulfan treatment group, and [redacted]% (95% CI: [redacted], [redacted]) in the treosulfan

treatment group. There was a statistically significant difference between the treatment groups (p=0.0343, adjusted for donor type as factor and risk group as stratum using Fine and Gray model). The HR was 0.63 (95% CI: 0.41, 0.97) in favour of treosulfan.

Table 21: Summary table of non-relapse mortality (FAS) - MC-FludT.14/L Trial II

	Busulfan (N=283)	**Treosulfan (N=268)**
Patients with event	56 (19.8%)	35 (13.1%)
Patients without event (censored) or with competing event	227 (80.2%)	233 (86.9%)
Censored	146 (51.6%)	171 (63.8%)
Relapse/Progression[a]	72 (25.4%)	61 (22.8%)
Primary Graft Failure[a]	1 (0.4%)	1 (0.4%)
Secondary Graft Failure[a]	8 (2.8%)	0 (0.0%)
Cumulative incidence at 12 months [%] (95% CI)	14.3 (10.2, 18.4)	10.5 (6.8, 14.2)
Cumulative incidence at 24 months [%] (95% CI)	20.4 (15.5, 25.2)	12.0 (8.0, 15.9)
Cumulative incidence at 36 months [%] (95% CI)	[redacted] ([redacted],[redacted])	[redacted] ([redacted],[redacted])
Hazard Ratio (treosulfan/busulfan)[b] (95% CI)	0.63 (0.41, 0.97)	
Adjusted p-value[b]	0.0343	
p-value[c]	0.0392	

Source: MC-FludT.14/L Trial II[60]

[a] Only if this event occurred first

[b] Adjusted for donor type as factor

In summary, the cumulative incidence of NRM at 24 months after alloHSCT was statistically significantly lower in the treosulfan treatment group compared with the busulfan treatment group (12.0% vs 20.4%; HR 0.63, 95% CI, 0.41, 0.97; adjusted p-value=0.0343). The cumulative incidence 36 months after alloHSCT was [redacted]% (95% CI; [redacted], [redacted]) in the treosulfan treatment group, and [redacted] (95% CI; [redacted], [redacted]) in the busulfan treatment group.

Transplantation-related mortality (TRM)

Transplantation-related mortality was assessed as a secondary endpoint of the study. The definition of “transplantation-related mortality” is provided in the Glossary.

A summary of TRM is presented in Table 22 for the FAS. At the time of analysis, 58 patients (20.5%) in the busulfan treatment group and 33 patients (12.3%) in the

treosulfan treatment group had died from a transplantation-related cause. The Kaplan-Meier estimate of TRM at 24 months was 24.1% (95% CI: 19.1%, 30.2%) for the busulfan treatment group, and 12.8% (95% CI: 9.2%, 17.7%) for the treosulfan treatment group. It is noteworthy that in the treosulfan group there was little change in TRM from 12 months to 24 months (TRM in treosulfan group at 12 months: 11.7%, 24 months: 12.8%). In the same timeframe in the busulfan treatment group, TRM increased from 16.2% to 24.1%. The difference between the treatment groups was statistically significant (p=0.0043, adjusted for donor type as factor, and risk group and centre as strata using Cox regression model). The HR was 0.52 (95% CI: 0.34, 0.82) in favour of treosulfan.

Table 22: Summary results of transplantation-related mortality (FAS) - MC-FludT.14/L Trial II

	Busulfan (N=283)	**Treosulfan (N=268)**
Patients with event	58 (20.5%)	33 (12.3%)
Patients without event	225 (79.5%)	235 (87.7%)
Transplantation-related mortality at 12 months[a] [%] (95% CI)	16.2 (12.2, 21.3)	11.7 (8.3, 16.3)
Transplantation-related mortality at 24 months[a] [%] (95% CI)	24.1 (19.1, 30.2)	12.8 (9.2, 17.7)
Hazard Ratio (treosulfan/busulfan)[b] (95% CI)	0.52 (0.34, 0.82)	
Adjusted p-value[b]	0.0043	
p-value[c]	0.0090	

Source: MC-FludT.14/L Trial II[60]

CI = confidence internal; N = number of patients.

[a] Based on Kaplan-Meier estimates.

[b] Adjusted for donor type as factor, and risk group and centre as strata using Cox regression model.

[c] Log-rank test.

The cumulative incidence of TRM was statistically significantly lower in the treosulfan treatment group compared with the busulfan treatment group (12.8% vs 24.1%, HR=0.52; 95% CI; 0.34, 0.82; adjusted p-value=0.0043). Of note, in the treosulfan group there was little change in TRM from 12 months (11.7%) to 24 months (12.8%), while in the same timeframe TRM increased from 16.2% to 24.1% in the busulfan treatment group. In addition, a statistically significantly lower cumulative incidence of TRM caused by infections was observed in the treosulfan treatment group compared with the busulfan treatment group (9.7% vs 17.2%; HR 0.57, 95% CI; 0.34, 0.97;

adjusted p-value=0.0371). Similarly, the cumulative incidence of TRM related to causes of death other than infections was statistically significantly lower in the treosulfan group than in the busulfan group (3.4% vs 8.4%; HR 0.42, 95% CI; 0.18, 0.97; adjusted p-value=0.0423).

Comparative evaluation of engraftment

A summary of graft failure is presented in Table 23 for the FAS. In the busulfan group, 1 patient (0.4%) experienced a primary graft failure, and 8 patients (2.9%) experienced a secondary graft failure. In the treosulfan group, 1 patient (0.4%) experienced a primary graft failure, and no patients experienced a secondary graft failure. No event of graft failure was reported during post-surveillance.

Table 23: Summary results of primary and secondary graft failure (FAS) - MC-FludT.14/L Trial II

	Busulfan	Treosulfan	Total
Full Analysis Set	N=283	N=268	N=551
Primary graft failure	1/283 (0.4%)	1/268 (0.4%)	2/551 (0.4%)
Secondary graft failure	8/279 (2.9%)	0/263 (0.0%)	8/542 (1.5%)

Source: MC-FludT.14/L Trial II[60]

Note: Rate of primary/secondary graft failure calculated as number of patients with graft failure by the number of patients at risk

- At risk for primary graft failure: Patients with HSCT
- At risk for secondary graft failure: Patients whose neutrophilic granulocytes engrafted after HSCT or were never below the required level

A rapid and sustained engraftment of neutrophils, leukocytes and platelets by day +28 after alloHSCT was observed in both treatment groups. The conditional cumulative incidence of reconstitution of granulopoiesis at day +28 after alloHSCT was 96.2% (95% CI; 93.4, 99.1) in the treosulfan group, and 96.8% (95% CI; 94.6, 99.1) in the busulfan treatment group. There was no statistically significant difference between the two treatment groups (p=0.4235). The median duration of neutropenia and leukopenia was statistically significantly longer in the treosulfan treatment group than in the busulfan treatment group (14.0 days vs 12.0 days; p<0.0001, and 14.0 days vs 13.0 days, adjusted p=0.0007, respectively).

In summary, within 24 months after alloHSCT, one graft failure was reported in the treosulfan group, in comparison with nine graft failures (one primary and eight

secondary) reported in the busulfan group. No event of graft failure was reported during post-surveillance.

Comparative evaluation of donor-type chimerism

Complete donor-type chimerism on Day +28 and Day +100 after HSCT was assessed as a secondary endpoint of the trial. Incidence of complete donor-type chimerism is presented in Table 24 for the FAS.

In the FAS, at the Day +28 visit, the incidence of complete donor-type chimerism was 83.3% (95% CI: 78.5%, 87.5%) in the busulfan treatment group, and 93.2% (95% CI: 89.4%, 95.9%) in the treosulfan group. At the Day +100 visit the incidences were 80.2% (95% CI: 74.9%, 84.9%) and 86.1% (95% CI: 81.2%, 90.1%) for the busulfan and treosulfan treatment groups, respectively. The differences at Day +28 (p=0.0159) and Day +100 (p=0.0381) were statistically significant in favour of treosulfan (both p-values for stratified Cochran-Mantel-Haenszel tests adjusted for donor type and risk group).

Table 24: Incidence of complete donor-type chimerism (FAS) - MC-FludT.14/L Trial II

	Busulfan	**Treosulfan**
Patients at risk at Day +28 visit[a]	**N=282**	**N=263**
Patients with complete chimerism at Day +28 visit	235 (83.3%)	245 (93.2%)
Patients without complete chimerism at Day +28 visit	41 (14.5%)	9 (3.4%)
Patients without information at Day +28 visit	6 (2.1%)	9 (3.4%)
Incidence of complete chimerism at Day +28 visit [%] (95% CI)	83.3 (78.5, 87.5)	93.2 (89.4, 95.9)
Odds ratio (95% CI)[b]	2.8083 (1.58, 5.01)	
Adjusted p-value[c]	0.0159	
p-valued	<.0001	
Patients at risk at Day +100 visit[a]	**N=263**	**N=252**
Patients with complete chimerism at Day +100 visit	211 (80.2%)	217 (86.1%)
Patients without complete chimerism at Day +100 visit	31 (11.8%)	24 (9.5%)
Patients without information at Day +100 visit	21 (8.0%)	11 (4.4%)
Incidence of complete chimerism at Day +100 visit [%] (95% CI)	80.2 (74.9, 84.9)	86.1 (81.2, 90.1)

	Busulfan	Treosulfan
Odds ratio (95% CI)[b]	1.5917 (0.99, 2.56)	
Adjusted p-value[c]	0.0381	
p-valued	0.1447	

Source: MC-FludT.14/L Trial II[60]

CI = confidence interval; N = number of patients.

[a] Patients are at risk if they have an examination at the Day +28 or Day +100 visit or if they have survived day +29 or day +107, respectively.

[b] Adjusted for donor type and risk group. Missing values accounted as 'No' for odds ratio calculation.

[c] Stratified Cochran-Mantel-Haenszel test adjusted for donor type and risk group.

[d] Crude Chi-Square test.

The incidence of complete donor-type chimerism was statistically significantly higher in the treosulfan treatment group compared with busulfan treatment group both at day +28 (93.2% vs 83.3%; adjusted p value=0.0159) and at day +100 (86.1% vs 80.2%; adjusted p-value=0.0381).

The incidence of complete donor-type chimerism, by donor type, risk group, disease, and age group, is presented in Table 25.

Table 25: Incidence of complete donor type chimerism by donor type, risk group, disease, and age group (FAS) - MC-FludT.14/L Trial II

		Busulfan		Treosulfan	
		n / N (%)	**95% CI**	**n / N (%)**	**95% CI**
Donor type					
MRD	Day +28	49 / 68 (72.1%)	(59.9,82.3)	52 / 60 (86.7%)	(75.4,94.1)
	Day +100	48 / 66 (72.7%)	(60.4,83.0)	46 / 57 (80.7%)	(68.1,90.0)
MUD	Day +28	186 / 214 (86.9%)	(81.6,91.1)	193 / 203 (95.1%)	(91.1,97.6)
	Day +100	163 / 197 (82.7%)	(76.7,87.7)	171 / 195 (87.7%)	(82.2,92.0)
Risk group					
Risk group I	Day +28	129 / 148 (87.2%)	(80.7,92.1)	115 / 124 (92.7%)	(86.7,96.6)
	Day +100	120 / 145 (82.8%)	(75.6,88.5)	112 / 121 (92.6%)	(86.3,96.5)
Risk group II	Day +28	106 / 134 (79.1%)	(71.2,85.6)	130 / 139 (93.5%)	(88.1,97.0)
	Day +100	91 / 118 (77.1%)	(68.5,84.3)	105 / 131 (80.2%)	(72.3,86.6)
Disease					
AML	Day +28	147 / 168 (87.5%)	(81.5,92.1)	171 / 181 (94.5%)	(90.1,97.3)
	Day +100	135 / 163 (82.8%)	(76.1,88.3)	152 / 175 (86.9%)	(80.9,91.5)
MDS	Day +28	88 / 114 (77.2%)	(68.4,84.5)	74 / 82 (90.2%)	(81.7,95.7)
	Day +100	76 / 100 (76.0%)	(66.4,84.0)	65 / 77 (84.4%)	(74.4,91.7)
Age group					

		Busulfan		Treosulfan	
< 50 years	Day +28	10 / 12 (83.3%)	(51.6,97.9)	15 / 16 (93.8%)	(69.8,99.8)
	Day +100	9 / 12 (75.0%)	(42.8,94.5)	11 / 16 (68.8%)	(41.3,89.0)
>= 50 years	Day +28	225 / 270 (83.3%)	(78.3,87.6)	230 / 247 (93.1%)	(89.2,95.9)
	Day +100	202 / 251 (80.5%)	(75.0,85.2)	206 / 236 (87.3%)	(82.4,91.3)

Source: MC-FludT.14/L Trial II[60]
Day +28 and Day +100 rates = No. of events (n) / No. at risk (N), where:
No. of events = No. of patients with complete chimerism
Patients are at risk if they have an examination at the Day +28 or Day +100 visit or if they have survived day +29 or day +107, respectively.

Interim analysis of secondary endpoints

For the secondary efficacy endpoints, results from this analysis on 570 patients are generally in line with those from the interim analysis on 476 patients.[72] Of note, this analysis showed a statistically significant advantage of treosulfan compared with busulfan for the incidence of NRM at 24 months (p=0.0343), and for complete donor type chimerism at Day +100 (p=0.0381).

Efficacy conclusions

This large, randomised Phase III study (n=570) showed that the treosulfan/fludarabine regimen can be successfully used as conditioning treatment prior to alloHSCT in adult patients with AML and MDS who are considered ineligible for standard MAC.

The treosulfan/fludarabine regimen was demonstrated to be statistically significantly non-inferior to the busulfan/fludarabine RIC regimen with regard to EFS at 24 months after alloHSCT (p=0.0000001). The treatment difference for EFS was statistically significantly in favour of treosulfan (p-value for testing difference vs busulfan=0.001), indicating a clinically relevant long-term advantage of treosulfan over busulfan.

Results from the primary endpoint analysis based on 570 patients are consistent with those from the interim analysis based on 476 patients.

The statistically significant advantage of treosulfan compared with busulfan was also demonstrated for the majority of the secondary endpoints: complete donor type chimerism (p=0.0381), OS (p=0.0037), NRM (p=0.0343), TRM (p=0.0043), GRFS (p=0.0087), and CRFS (p=0.0030).

The cumulative incidence of relapse/progression at 24 months was comparable between the two regimens.

The robustness of these results was confirmed by subgroup and sensitivity analyses.

In the post-surveillance evaluation, follow-up data for EFS, OS, relapse/progression, and NRM showed a continued clinically relevant long-term advantage of treosulfan compared with busulfan for these clinically meaningful endpoints.

B.2.7 Subgroup analysis

In MC-FludT.14/L Trial II subgroup analyses were implemented in order to investigate the consistency of the study results. The specific subgroups were identified at the stage of development of the SAP[76] based on potential prognostic factors discussed in the literature. All subgroup analyses were of exploratory nature only.

The following subgroups were considered: centre, risk group, donor type (MUD vs MRD), combination of donor type and risk group, disease (AML vs MDS), remission status at study entry (CR1 vs > CR1, for AML only), disease status at study entry (treated vs untreated, for MDS only), age group (<50 years vs ≥50 years), and HCT-CI Score (≤2 vs >2).

The subgroup analyses for EFS was presented as a Forest plot earlier (see Figure 5).

Forest plots for OS at 24 months by prognostic factors and combinations of prognostic factors is presented for the FAS (Figure 6). In the disease subgroups, OS for AML patients (FAS) was [redacted]% in the busulfan treatment group ([redacted] patients), and [redacted]% in the treosulfan treatment group ([redacted] patients). For MDS patients, OS was [redacted]% in the busulfan treatment group ([redacted] patients), and [redacted]% in the treosulfan treatment group ([redacted] patients). In the AML remission status subgroup, OS for patients in CR1 was [redacted]% in the busulfan treatment group (144 patients), and [redacted]% for the treosulfan treatment group ([redacted] patients). The results of the subgroup analyses were consistent with the analysis of the total population. The HRs for MRD, MRD risk group II, and MDS risk group I were in favour of busulfan (HR > 1.0). All other HRs

for each subgroup were in favour of treosulfan (HR < 1.0). As the CIs for each prognostic factor or combination were widely overlapping, no differential effects could be identified.[60]

Within this trial, comparable results in favour of treosulfan have also been seen with longer follow-up at 36 months, showing that the results are durable.

Figure 6: Forest plot for OS by prognostic factors with 24 month event rates (FAS) – MC-FludT.14/L Trial II

Source: MC-FludT.14/L Trial II[60]

B.2.8 Meta-analysis

Based on the systematic literature review (Appendix D), a feasibility analysis was performed to assess whether network meta-analyses could be performed to provide

comparative evidence for treosulfan + fludarabine (TREO/FLU) versus other conditioning regimens of interest based on the NICE scope (Appendix L).

Of the 28 references identified in the clinical SLR, four RCTs[59,85–89] including the pivotal phase III study, were deemed relevant for inclusion in the analysis in that they included regimens of interest (i.e. TREO/FLU, busulfan/fludarabine [BU/FLU] and busulfan/cyclophosphamide [BU/Cy]) and reported data for the efficacy endpoints reported for the pivotal trial (i.e. OS, EFS, NRM, relapse rate (RR) or GvHD).

The feasibility assessment findings indicate that it is not possible to perform a network meta-analysis for any of the efficacy endpoints reported in the pivotal phase III trial.

The feasibility assessment can be found in (Appendix L) and details for each of the pivotal trial efficacy endpoints are below.

B.2.8.1 Overall survival (OS)

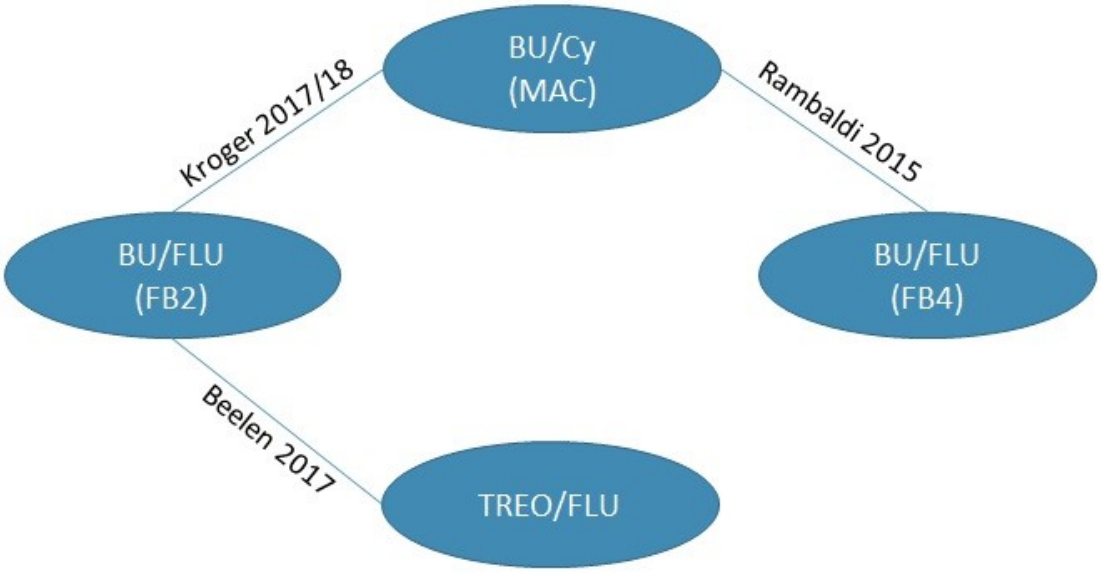
OS were available from four studies for 1, 2 and 5 year time points as shown in Table 26.

Table 26: Studies reporting data for overall survival

Study	1 year	2 years	5 years
Beelen et al 2017/19[59,85] (MC-FludT.14/L Trial II)	X	✓	X
Kroger et al 2017/18[87,90]	X	✓	✓
Liu et al 2013[88]	X	X	✓
Rambaldi et al 2015[89]	✓	✓	✓

The results indicated that a network meta-analysis is not feasible for any of the time points (Figure 7).

Figure 7: Comparisons for OS at 2 years: AML/MDS



AML, Acute myeloid leukaemia; BU/Cy, busulfan/cyclophosphamide; BU/FLU, busulfan/fludarabine; MAC, Myeloablative conditioning; MDS, Myelodysplastic syndrome; OS, Overall survival; TREO/FLU, treosulfan/fludarabine

B.2.8.2 Event-free survival (EFS)

EFS was reported in the phase III TREO/FLU MC-FludT.14/L Trial II and was the primary endpoint. However, this endpoint was not reported in any of the other trials and hence cannot be compared with results for comparator regimens.

B.2.8.3 Non-relapse mortality (NRM)

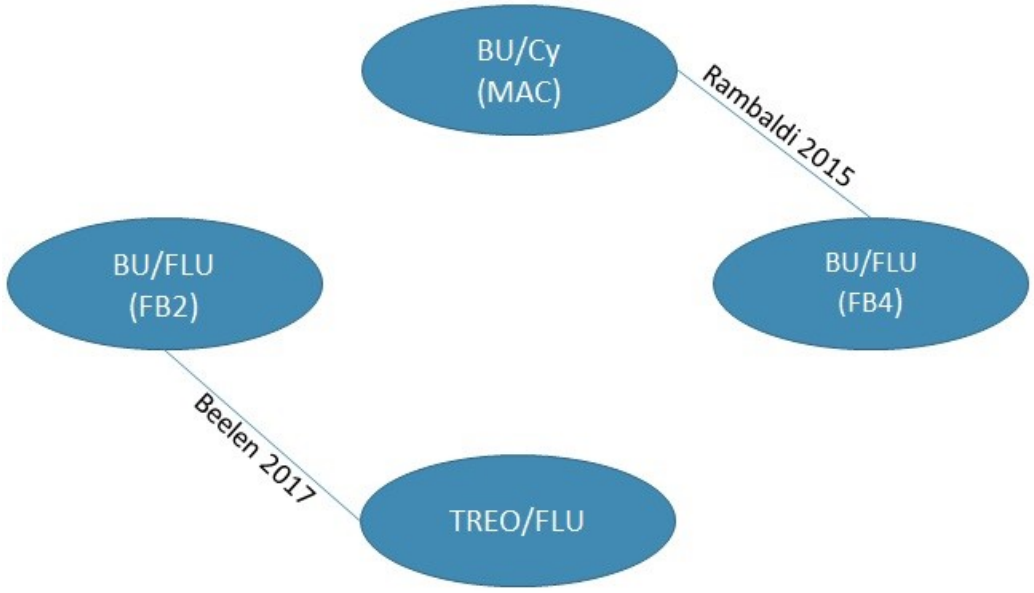
NRM data were available from three studies at 2 years and 5 years (Table 27).

Table 27: Studies reporting data for NRM

Study	2 years	5 years
Beelen et al 2017/19[59,85] (MC-FludT.14/L Trial II)	✓	X
Kroger et al 2017/18[87,90]	X	✓
Rambaldi et al 2015[89]	✓	✓

A network meta-analysis is not feasible at either time point (Figure 8)

Figure 8: Comparisons for NRM at 2 years: AML/MDS



AML, Acute myeloid leukaemia; BU/Cy, busulfan/cyclophosphamide; BU/FLU, busulfan/fludarabine; MAC, Myeloablative conditioning; MDS, Myelodysplastic syndrome; NRM, Non-relapse mortality; OS, Overall survival; TREO/FLU, treosulfan/fludarabine

B.2.8.4 Relapse rate

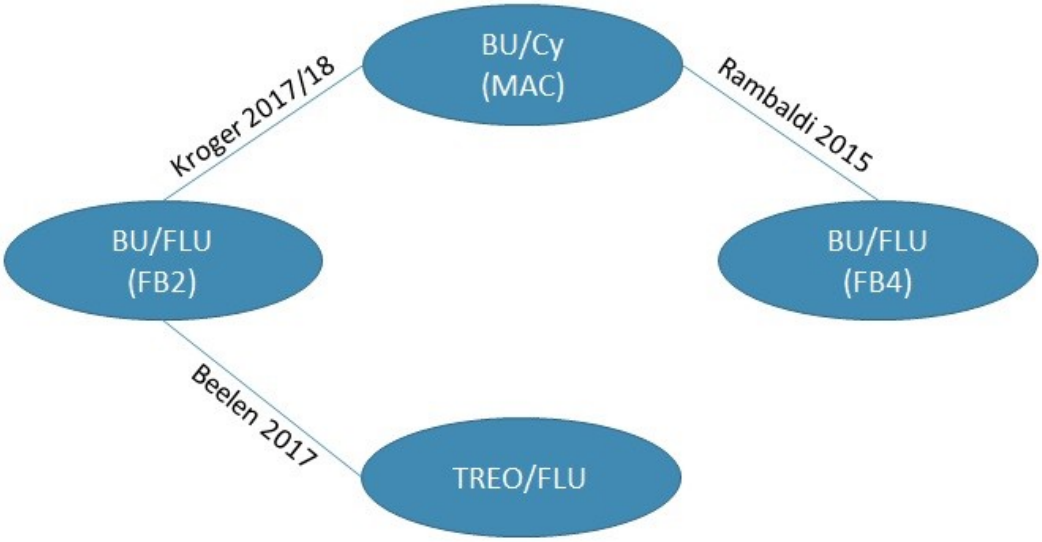
Relapse rate was reported in four studies for 1, 2 and 5 year time points (Table 28).

Table 28: Studies reporting data for relapse rate

Study	1 year	2 years	5 years
Beelen et al 2017/19[59,85] (MC-FludT.14/L Trial II)	X	✓	X
Kroger et al 2017/18[87,90]	X	✓	✓
Liu et al 2013[88]	X	X	✓
Rambaldi et al 2015[89]	✓	✓	✓

No network meta-analyses versus TREO/FLU are possible at any time point. (Figure 9).

Figure 9: Comparisons for relapse rate at 2 years: AML/MDS



AML, Acute myeloid leukaemia; BU/Cy, busulfan/cyclophosphamide; BU/FLU, busulfan/fludarabine; MAC, Myeloablative conditioning; MDS, Myelodysplastic syndrome; TREO/FLU, treosulfan/fludarabine

B.2.8.5 Chronic graft versus host disease (GvHD)

The incidence of chronic GvHD was reported in four studies at 1 and 2 years (Table 29).

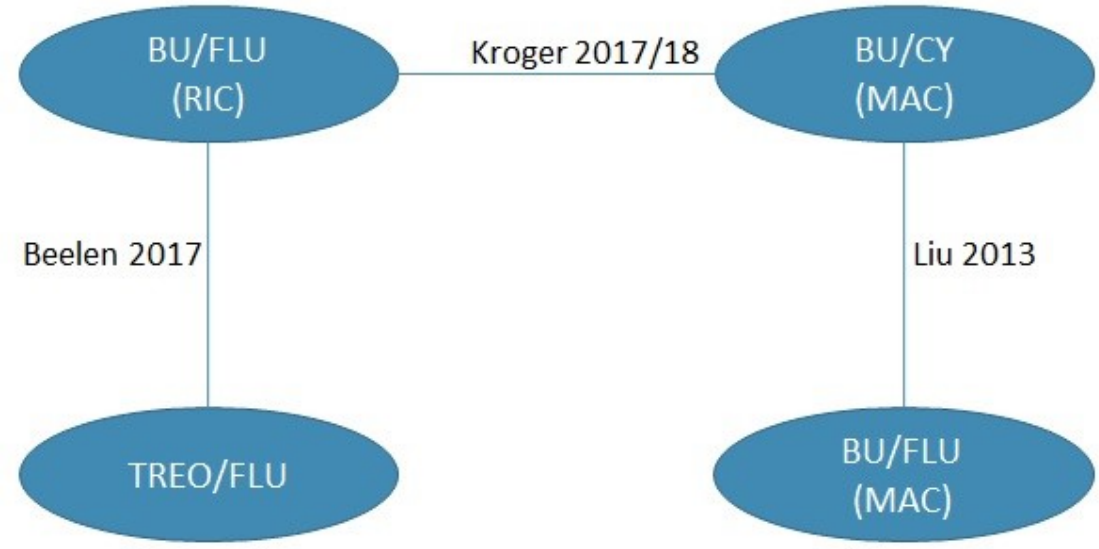
Table 29: Studies reporting data for chronic GvHD

Study	1 year	2 years
Beelen et al 2017/19[59,85] (MC-FludT.14/L Trial II)	X	✓
Kroger et al 2017/18[87,90]	X	✓
Liu et al 2013[88]	X	✓
Rambaldi et al 2015[89]	✓	X

cGvHD, chronic graft-versus-host disease

A network meta-analysis is not feasible at either 1 year or 2 years (Figure 10).

Figure 10: Comparisons for chronic GvHD at 2 years: AML/MDS



AML, Acute myeloid leukaemia; BU/Cy, busulfan/cyclophosphamide; BU/FLU, busulfan/fludarabine; GvHD, Graft-versus-host disease; MAC, Myeloablative conditioning; MDS, Myelodysplastic syndrome; RIC, Reduced intensity conditioning; TREO/FLU, treosulfan/fludarabine

In summary, the feasibility assessment considered the data available from the pivotal TREO/FLU MC-FludT.14/L Trial II and the three other RCTs identified in the SLR and concluded that it is not possible to perform a network meta-analysis for any of the efficacy endpoints reported in the pivotal phase III trial.

B.2.9 Indirect and mixed treatment comparisons

Summary of clinical evidence for comparator therapies

- A feasibility assessment for the completion of indirect and mixed treatment comparisons to provide comparative evidence for TREO/FLU versus other conditioning regimens of interest based on the NICE scope (Appendix L).
- Of the 28 references identified in the clinical SLR, four RCTs,[59,85–89] including the pivotal phase III study (MC-FludT.14/L Trial II) were deemed relevant for inclusion in the analysis as they included regimens of interest (TREO/FLU, BU/FLU and busulfan/cyclophosphamide (BU/Cy)) and reported data for the efficacy endpoints reported for the pivotal trial (OS, EFS, NRM, RR or GvHD).
- Some indirect comparisons would be possible (for TREO/FLU versus BU/Cy and BU/FLU (MAC) at 2 years for OS, RR and the incidence of GvHD). However, these outcomes are unlikely to provide sufficient, reliable and relevant comparative data for inclusion in the economic assessment of TREO/FLU as a conditioning regimen for patients

undergoing HSCT as treatment for malignant disease.

A feasibility assessment for the completion of indirect and mixed treatment comparisons was completed based on the systematic literature review. The feasibility assessment considered whether indirect comparisons could be performed to provide comparative evidence for TREO/FLU versus other conditioning regimens of interest based on the NICE scope (Appendix L).

As noted for the meta-analysis feasibility assessment, of the 28 references identified in the clinical SLR, four RCTs,[59,85–89] including the pivotal phase III study, were deemed relevant for inclusion in the analysis in that they included regimens of interest (i.e. TREO/FLU, BU/FLU and busulfan/cyclophosphamide (BU/Cy)) and reported data for the efficacy endpoints reported for the pivotal trial (i.e. OS, EFS, NRM, RR or GvHD). The full feasibility assessment report is in Appendix L and details for the efficacy endpoints from the pivotal trial are provided below.

B.2.9.1 Overall survival (OS)

OS were available from four studies for 1, 2 and 5 year time points as shown in Table 26. These results indicate that an indirect comparisons vs TREO/FLU may be feasible vs BU/Cy and BU/FLU (MAC) at 2 years.

B.2.9.2 Event-free survival (EFS)

EFS was reported in the phase III MC-FludT.14/L Trial II and was the primary endpoint. However, this endpoint was not reported in any of the other trials and hence cannot be compared with results for comparator regimens.

B.2.9.3 Non-relapse mortality (NRM)

NRM data were available from three studies at 2 years and 5 years (Table 27). However, an indirect comparisons versus TREO/FLU is not possible at either time point (Figure 8).

B.2.9.4 Relapse rate

Relapse rate was reported in four studies for 1, 2 and 5 year time points (Table 28). At 2 years, an indirect comparisons versus TREO/FLU is possible for BU/Cy and BU/FLU (MAC) (Figure 9). However, relapse rate is not included in the economic model and thus would not provide any additional data to inform the economic model.

B.2.9.5 Chronic graft versus host disease (GvHD)

The incidence of chronic GvHD was reported in four studies at 1 and 2 years (Table 29). Indirect comparisons for TREO/FLU versus BU/Cy and BU/Flu (MAC) are feasible at 2 years (Figure 10).

In summary, indirect comparisons of TREO/FLU may be possible versus BU/Cy and BU/FLU (MAC) at 2 years for OS, RR and the incidence of chronic GvHD, based on the results from the pivotal phase III trial, Kroger et al 2017/18[87,90] and Rambaldi et al 2015.[89] However, while the pivotal phase III trial and Kroger et al 2017/18[87,90] involved the same BU/FLU regimen, there are non-disease related differences in the patients involved in Kroger et al 2017/18[87,90] and the pivotal phase III trial which may limit the validity of an indirect comparison between the two trials, namely median age (60 years vs 50 years) and the proportion of patients receiving matched unrelated alloHSCT (76% vs 37%)]. The Kroger et al 2017/18[87,90] and Rambaldi et al 2015[89] studies both involved the same BU/Cy regimen but again there are some differences in the patient populations between the two studies which may affect the validity of indirect comparisons; the proportion of patients receiving unrelated alloHSCT was 37% in Kroger et al and 55% in Rambaldi et al. Furthermore, the endpoints for which indirect comparisons could be compared are not those that are the most appropriate for comparing the efficacy of the different conditioning regimens. Although OS is included in the model and is a relevant clinical outcome, it is affected by many different factors including patient risk factors and subsequent treatments on relapse. RR and GvHD are relevant clinical outcomes for comparing different conditioning regimens but are either not included, or are not key drivers of, the economic model and would not provide sufficient data to inform the totality of the economic model which includes additional clinical variables such as EFS.

In conclusion, while indirect comparisons may be possible for TREO/FLU versus BU/Cy and BU/FLU (MAC) at 2 years for OS, RR and the incidence of GvHD, these outcomes are unlikely to provide sufficient, reliable and relevant comparative data for inclusion in the economic assessment of TREO/FLU as a conditioning regimen for patients undergoing HSCT as treatment for malignant disease.

B.2.10 MC-FludT.14/L Trial II adverse reactions

Summary of MC-FludT.14/L Trial II safety analysis

- Safety results from this study showed that the tolerability profiles of the treosulfan-based regimen and the busulfan-based regimen were comparable.
- Overall, the frequency of treatment emergent adverse event (TEAEs) (treosulfan, 92.6%; busulfan, 96.1%) and serious adverse events (SAEs) (treosulfan, 8.5%; busulfan, 7.1%) were comparable in the two treatment groups.[60]
- The severity of TEAEs was similar in the two treatment groups:[60]
  - The majority of patients reported events of CTCAE Grade III or lower (treosulfan, 83.4%; busulfan, 90.1%).
  - TEAEs of CTCAE Grade III or higher were reported by approximately half of the patients in both treatment groups (treosulfan, 54.8%; busulfan, 53.4%).
  - Drug-related TEAEs of Grade III or higher were reported by approximately a third of patients in both treatment groups (treosulfan, 26.7%; busulfan, 29.0%). Investigations (treosulfan, 10.4%; busulfan, 9.5%) and gastrointestinal disorders (treosulfan, 9.6%; busulfan, 11.3%) were the most commonly reported TEAEs of Grade III or higher.
- A numerical advantage in favour of treosulfan was observed for the incidence of mucositis of CTCAE Grade III/IV (treosulfan, 5.9%; busulfan, 7.4%) and nausea (treosulfan, 3.0%; busulfan, 6.0%); however, these differences were not statistically significant.[60]
- The overall incidence of deaths during the study period was 32.4% (treosulfan, 26.7%; busulfan, 37.8%), with relapse/progression and transplantation being the most common causes of death.[60]
- The cumulative incidence of acute and chronic GvHD was similar in the two treatment groups.[60]
  - A numerical advantage in favour of treosulfan compared with busulfan was observed in both Grade I to IV acute GvHD (52.8% vs 57.2%) and Grade III/IV acute GvHD (6.4% vs 8.1%); however, these differences were not statistically significant.

- A non-statistically significant numerical advantage in favour of treosulfan was observed for extensive chronic GvHD at 24 months (treosulfan, 19.8%; busulfan 28.6%).

- Patients' functional ability, general well-being and ability to perform daily life activities following transplantation, as measured by Karnofsky index, were in favour of treosulfan. The "time to deterioration of Karnofsky Performance Score [KPS] by ≥20 points" and "deterioration of the KPS to <60 points" were significantly in favour of the treosulfan treatment group (p=0.0261 and p=0.0130, respectively).[60]
- Results from this safety analysis based on 570 patients are consistent with those from the second interim analysis based on 476 patient.[60,72]

The secondary objectives of this study were the comparative evaluation of the following safety endpoints:

- Comparative evaluation of incidence of CTC grade III/IV mucositis between Day -6 and Day +28
- Comparative evaluation of cumulative incidence of acute and chronic GvHD within 2 years after transplantation
- Comparative evaluation of incidence of other CTC grade III/IV adverse events between Day -6 and Day +28

Other safety endpoints evaluated included:

- Incidence of treatment related adverse events (TEAEs) between day-6 and day+28
- Incidence of deaths during the study period
- Changes of KPS – a measure of patient functional ability, status of general well-being and activities of daily life – from baseline throughout the study.

Adverse event / reaction collection

Adverse events according to the definition given the Clinical Trial Protocol were to be reported between Day -6 and Day +28 of the trial. According to this Clinical Trial Protocol, the following events were not to be documented on the AE-pages:

- any changes in blood counts and differential blood counts
- symptoms definitely caused by episodes of GvHD were documented separately

- all other laboratory values out of normal range unless a clinical relevance was assumed by the investigator.

The occurrence of GvHD either acute (until Day +100) or chronic (day +101 until 2 years after transplantation) was to be documented throughout the entire study period, which was 2 years after transplantation at most.

An AE was classified as related to IMP ("drug related") if the relationship was classified as "possibly related" or "related" by the Investigator.

After Day +28, only SAEs with a suspected relationship to the investigational medicinal product (IMP) were documented.

B.2.10.1 Treatment-related adverse events (TEAEs) and deaths

Overall, the frequency of TEAEs was comparable in the two treatment groups (Table 30). As expected with an intervention as severe as alloHSCT, TEAEs were reported by the majority of patients in both treatment groups (treosulfan, 92.6%; busulfan, 96.1%). SAEs were reported by 8.5% of patients in the treosulfan treatment group and 7.1% of patients in the busulfan treatment group. Drug-related SAEs were reported by a small proportion of patients in both treatment groups (treosulfan, 3.3%; busulfan, 3.2%). The severity of AEs, as measured by the maximum CTCAE Grade of AEs, was similar in the two treatment groups, with the majority patients reporting events of CTCAE Grade III or lower (treosulfan, 83.4%; busulfan, 90.1%).

Table 30: Summary of treatment emergent adverse events (Safety Analysis Set; SAS) – MC FludT.14/L Trial II

	Busulfan (N=283)	**Treosulfan (N=270)**	**Total (N=553)**
Any AE, n (%)			
Patients with any AE	272 (96.1)	250 (92.6)	522 (94.4)
Patients with AEs of at least CTCAE Grade III	151 (53.4)	148 (54.8)	299 (54.1)
Drug-related AEs, n (%)			
Patients with any drug-related AEs	192 (67.8)	170 (63.0)	362 (65.5)
Patients with drug-related AEs of at least CTCAE Grade III	82 (29.0)	72 (26.7)	150 (27.8)
SAE, n (%)			
Patients with any SAE	20 (7.1)	23 (8.5)	43 (7.8)

	Busulfan (N=283)	Treosulfan (N=270)	Total (N=553)
Resulting in death	6 (2.1)	8 (3.0)	14 (2.5)
Life-threatening	8 (2.8)	13 (4.8)	21 (3.8)
Hospitalisation or prolongation of hospitalisation	9 (3.2)	8 (3.0)	17 (3.1)
Drug-related SAEs, n (%)			
Patients with any drug related SAE	9 (3.2%)	9 (3.3%)	18 (3.3%)
Maximum CTCAE grade of adverse events [n (%)]			
Patients with AEs of a maximum CTCAE grade I	46 (16.3%)	41 (15.2%)	87 (15.7%)
Patients with AEs of a maximum CTCAE grade II	75 (26.5%)	61 (22.6%)	136 (24.6%)
Patients with AEs of a maximum CTCAE grade III	134 (47.3%)	123 (45.6%)	257 (46.5%)
Patients with AEs of a maximum CTCAE grade IV	14 (4.9%)	18 (6.7%)	32 (5.8%)
Patients with AEs of a maximum CTCAE grade V	3 (1.1%)	7 (2.6%)	10 (1.8%)

Source: MC-FludT.14/L Trial II[60]
CTCAE = Common Terminology Criteria for Adverse Events; N = number of patients; n = number of patients in category.

The frequency of TEAE of Grade III or higher by CTCAE System Organ Class and Term occurring in ≥5% of patients in either treatment group is summarised in Table 31. TEAEs of CTCAE Grade III or higher were reported by 53.4% of patients in the busulfan treatment group, and 54.8% of patients in the treosulfan treatment group. More patients in the busulfan group than in the treosulfan group reported an event in the class gastrointestinal disorders (15.5% vs 12.2% of patients). Conversely, more patients in the treosulfan group than the busulfan group reported an event in the class infections and infestations (15.2% vs 9.2% of patients). A numerical advantage in favour of treosulfan was observed for mucositis oral CTCAE Grade III or higher (treosulfan, 5.9%; busulfan, 7.4%) and nausea (treosulfan, 3.0%; busulfan 6.0%); however, the differences between groups were not statistically significant. Data for the other TEAEs classes were broadly similar for the two treatment groups.

Drug-related TEAEs of Grade III or higher were reported by approximately a third of patients in both treatment groups (treosulfan, 26.7%; busulfan, 29.0%). Such events were reported by ≥5% of patients in either treatment group for three TEAEs classes: investigations (treosulfan, 10.4%; busulfan, 9.5%) and gastrointestinal disorders (treosulfan, 11.3%; busulfan, 9.6%).

Table 31: Frequency of patients with at least CTCAE Grade III treatment emergent adverse events by CTCAE System Organ Class and Term occurring in at least 5% of patients in either treatment group (SAS) – MC-FludT.14/L Trial II

CTCAE System Organ Class/Term	Busulfan (N=283)	Treosulfan (N=270)	Total (N=553)
Patients with any event	151 (53.4%)	148 (54.8%)	299 (54.1%)
Gastrointestinal disorders			
Any event	44 (15.5%)	33 (12.2%)	77 (13.9%)
Mucositis oral	21 (7.4%)	16 (5.9%)	37 (6.7%)
Nausea	17 (6.0%)	8 (3.0%)	25 (4.5%)
Investigations			
Any event	38 (13.4%)	39 (14.4%)	77 (13.9%)
Gamma-glutamyl transferase increased	25 (8.8%)	12 (4.4%)	37 (6.7%)
Alanine aminotransferase increased	9 (3.2%)	14 (5.2%)	23 (4.2%)
Blood and lymphatic system disorders			
Any event	31 (11.0%)	40 (14.8%)	71 (12.8%)
Febrile neutropenia	31 (11.0%)	40 (14.8%)	71 (12.8%)
Infections and infestations			
Any event	26 (9.2%)	41 (15.2%)	67 (12.1%)
Vascular disorders			
Any event	36 (12.7%)	27 (10.0%)	63 (11.4%)
Hypertension	27 (9.5%)	21 (7.8%)	48 (8.7%)
Metabolism and nutrition disorders			
Any event	16 (5.7%)	24 (8.9%)	40 (7.2%)

Source: MC-FludT.14/L Trial II[60]
CTCAE = Common Terminology Criteria for Adverse Events; N = number of patients.

These results are consistent with those from the second interim analysis of this study based on 476 patients.[72]

B.2.10.2 Acute and chronic GvHD

B.2.10.2.1 Incidence of GvHD-free and relapse/progression-free survival (GRFS) within 2 years of HSCT

The incidence of GRFS within 2 years of HSCT was assessed as a secondary endpoint of the trial. A summary of GRFS data is presented in Table 32 for the FAS.

Table 32: Summary results of GvHD-free and relapse/progression-free survival chimerism (FAS) - MC-FludT.14/L Trial II

	Busulfan (N=283)	**Treosulfan (N=268)**
Patients with event	**169 (59.7%)**	**130 (48.5%)**
Death[a]	30 (10.6%)	21 (7.8%)
Relapse/Progression[a]	64 (22.6%)	54 (20.1%)
Acute GvHD ≥ Grade III[a]	23 (8.1%)	17 (6.3%)
Extensive chronic GvHD[a]	52 (18.4%)	38 (14.2%)
Patients without event	**114 (40.3%)**	**138 (51.5%)**
GvHD-free and relapse/progression-free survival at 12 months[b] [%] (95% CI)	49.8 (43.7, 55.5)	56.9 (50.7, 62.6)
GvHD-free and relapse/progression-free survival at 24 months[b] [%] (95% CI)	37.1 (31.1, 43.1)	50.3 (43.9, 56.3)
Hazard Ratio (treosulfan/busulfan)[c] (95% CI)	0.73 (0.57, 0.92)	
Adjusted p-value[c]	0.0087	
p-valued	0.0115	

Source: MC-FludT.14/L Trial II[60]
CI = confidence interval; GvHD = graft-versus-host disease.
Note: GvHD-free defined as no acute GvHD of at least grade III and no extensive chronic GvHD.
[a] Only if this event occurred first.
[b] Based on Kaplan-Meier estimates.

At the end of the trial, 40.3% of patients in the busulfan treatment group and 51.5% of patients in the treosulfan treatment group were alive and had not experienced GvHD or relapse/progression. The Kaplan-Meier estimates of GRFS at 24 months were 37.1% (95% CI: 31.1%, 43.1%) for the busulfan treatment group, and 50.3% (95% CI: 43.9%, 56.3%) for the treosulfan treatment group. The difference between the treatment groups was statistically significant (p=0.0087, adjusted for donor type as factor, and risk group and centre as strata using Cox regression model). The HR was 0.73 (95% CI: 0.57, 0.92) in favour of treosulfan.

Chronic GvHD-free and relapse/progression-free survival (CRFS) was also statistically significantly higher in the treosulfan treatment group (51.4% vs 37.2%; HR 0.70, 95% CI; 0.55, 0.88; adjusted p-value=0.0030).

In summary GvHD-free and relapse/progression-free survival was statistically significantly higher in the treosulfan treatment group compared with the busulfan

treatment group (50.3% vs 37.1%; HR 0.73, 95% CI; 0.57, 0.92; adjusted p-value=0.0087).

B.2.10.2.2 Acute GvHD

Acute GvHD was classified as GvHD up until Day +100 after HSCT.

In the FAS, the incidence of acute GvHD grade I-IV was numerically lower in the treosulfan treatment group compared to the busulfan treatment group, but the difference was not statistically significant (p=0.2038, based on the test of Gray). Acute GvHD grade I-IV was reported by 57.2% of busulfan patients, and 52.6% of treosulfan patients. The cumulative incidence at 100 days was 57.2 (95% CI: 51.5, 63.0) for the busulfan treatment group, and 52.8% (95% CI: 46.8, 58.8) for the treosulfan treatment group. The HR was 0.87 (95% CI: 0.69, 1.08) in favour of treosulfan.

Acute GvHD grade II-IV was reported by 22.3% of patients in the busulfan treatment group, and 19.4% of patients in the treosulfan treatment group. The cumulative incidence at 100 days was 22.3% (95% CI: 17.4, 27.1) for the busulfan treatment group, and 19.5% (95% CI: 14.7, 24.2) for the treosulfan treatment group. The difference between the treatment groups was not statistically significant (p=0.4161, based on test of Gray). The HR was 0.86 (95% CI: 0.60, 1.24) in favour of treosulfan.

Acute GvHD of grade III to IV was reported by 8.1% of patients in the busulfan treatment group and 6.3% of patients in the treosulfan treatment group (Table 33). The cumulative incidence at 100 days was 8.1% (95% CI: 4.9, 11.3) in the busulfan treatment group and 6.4% (95% CI: 3.4, 9.3) in the treosulfan treatment group. The difference between the treatment groups was not statistically significant (p=0.4267, based on test of Gray).

Table 33: Summary table of acute GvHD (FAS) – MC-FludT.14/L Trial II

	Busulfan (N=283)	**Treosulfan (N=268)**
Acute GvHD grade I-IV		
Patients with event	162 (57.2%)	141 (52.6%)
Patients without event (censored) or with competing event	121 (42.8%)	127 (47.4%)
Censored	97 (34.3%)	101 (37.7%)

	Busulfan (N=283)	Treosulfan (N=268)
Death[a]	4 (1.4%)	11 (4.1%)
Relapse/Progression[a]	17 (6.0%)	14 (5.2%)
Primary Graft Failure[a]	1 (0.4%)	1 (0.4%)
Secondary Graft Failure[a]	2 (0.7%)	0 (0.0%)
Cumulative incidence at 14 days [%] (95% CI)	14.1 (10.1, 18.2)	10.5 (6.8, 14.2)
Cumulative incidence at 28 days [%] (95% CI)	36.7 (31.1, 42.4)	30.0 (24.5, 35.5)
Cumulative incidence at 100 days [%] (95% CI)	57.2 (51.5, 63.0)	52.8 (46.8, 58.8)
Hazard Ratio (treosulfan/busulfan) (95% CI)	0.87 (0.69, 1.08)	
p-value[b]	0.2038	
Acute GvHD grade III-IV		
Patients with event	23 (8.1%)	17 (6.3%)
Patients without event (censored) or with competing event	260 (91.9%)	251 (93.7%)
Censored	215 (76.0%)	214 (79.9%)
Death[a]	7 (2.5%)	11 (4.1%)
Relapse/Progression[a]	34 (12.0%)	25 (9.3%)
Primary Graft Failure[a]	1 (0.4%)	1 (0.4%)
Secondary Graft Failure[a]	3 (1.1%)	0 (0.0%)
Cumulative incidence at 14 days [%] (95% CI)	0.7 (0.0, 1.7)	0.7 (0.0, 1.8)
Cumulative incidence at 28 days [%] (95% CI)	2.8 (0.9, 4.8)	1.1 (0.0, 2.4)
Cumulative incidence at 100 days [%] (95% CI)	8.1 (4.9, 11.3)	6.4 (3.4, 9.3)
Hazard Ratio (treosulfan/busulfan) (95% CI)	0.78 (0.42, 1.45)	
p-value[b]	0.4267	

Source: MC-FludT.14/L Trial II[60]

CI = confidence interval; GvHD = graft-versus-host disease; N = number of patients.

[a] Only if this event occurred first.

[b] Based on test of Gray.

B.2.10.2.3 Chronic GvHD

Chronic GvHD was classified as GvHD after Day +100. A summary of chronic GvHD is presented in Table 34. The cumulative incidence of chronic GvHD at 24 months was similar in the two treatment groups (treosulfan, 61.7%; busulfan 60.3%). A non-statistically significant numerical advantage in favour of treosulfan was observed for extensive chronic GvHD at 24 months compared with busulfan (19.8% vs 28.6%; p=0.0750) (Table 34).

Table 34: Summary table of chronic GvHD (FAS) – MC-FludT.14/L Trial II

	Busulfan (N=283)	Treosulfan (N=268)
Chronic GvHD		
Patients at risk[a]	232	229
Patients with event	138 (59.5%)	138 (60.3%)
Patients without event (censored) or with competing event	94 (40.5%)	91 (39.7%)
Censored	58 (25.0%)	59 (25.8%)
Death[b]	9 (3.9%)	9 (3.9%)
Relapse/Progression[b]	22 (9.5%)	23 (10.0%)
Primary Graft Failure[b]	0 (0.0%)	0 (0.0%)
Secondary Graft Failure[b]	5 (2.2%)	0 (0.0%)
Cumulative incidence at 6 months [%] (95% CI)	41.9 (35.5, 48.2)	40.3 (34.0, 46.7)
Cumulative incidence at 12 months [%] (95% CI)	55.1 (48.7, 61.5)	54.8 (48.4, 61.3)
Cumulative incidence at 24 months [%] (95% CI)	60.3 (53.8, 66.7)	61.7 (55.1, 68.3)
Hazard Ratio (treosulfan/busulfan) (95% CI)	1.00(0.79, 1.27)	
p-value[c]	0.9964	
Extensive chronic GvHD		
Patients at risk[a]	232	229
Patients with event	62 (26.7%)	45 (19.7%)
Patients without event (censored) or with competing event	170 (73.3%)	184 (80.3%)
Censored	110 (47.4%)	140 (61.1%)
Death[b]	24 (10.3%)	13 (5.7%)
Relapse/Progression[b]	31 (13.4%)	31 (13.5%)
Primary Graft Failure[b]	0 (0.0%)	0 (0.0%)
Secondary Graft Failure[b]	5 (2.2%)	0 (0.0%)
Cumulative incidence at 6 months [%] (95% CI)	13.8 (9.4, 18.2)	11.4 (7.3, 15.5)
Cumulative incidence at 12 months [%] (95% CI)	20.0 (14.8, 25.2)	16.7 (11.9, 21.6)
Cumulative incidence at 24 months [%] (95% CI)	28.6 (22.5, 34.7)	19.8 (14.5, 25.1)
Hazard Ratio (treosulfan/busulfan) (95% CI)	0.71 (0.48, 1.04)	
p-value[c]	0.0750	

Source: MC-FludT.14/L Trial II[60]

CI = confidence interval; GvHD = graft-versus-host disease; N = number of patients.

[a] Patients are at risk if they have survived 100 days after end of HSCT without relapse and graft failure.

[b] Only if this event occurred first

[c] Based on test of Gray

These results are consistent with those from the second interim analysis based on 476 patients.[72]

B.2.10.3 Karnofsky Performance Score

Karnofsky Performance Score (KPS), a measure of patient functional ability, status of general well-being and activities of daily life, was recorded throughout the trial.

At baseline, median KPS was 90 in both treatment groups. During the trial, the median KPS for patients in the busulfan treatment group fell from 90 at day 0 to 80 at day +100 and month 6 and month 9, then increased back to 90 at month 12 until the end of the trial. For patients in the treosulfan treatment group, the median KPS was 80 at day +100, and 90 for all other evaluated time points. Accordingly, the median KPS score was in favour of the treosulfan treatment group at 2 of 9 post HSCT time points.

B.2.10.4 Deterioration of KPS by at least 20 points

The time-dependent rate of patients experiencing a specific clinically meaningful deterioration was also analysed. The Kaplan-Meier estimate at 24 months for patients showing a deterioration ≥20 points was 60.0% (95% CI; 53.5%, 66.5%) for the busulfan treatment group, and 50.3% (95% CI; 44.1%, 56.9%) for the treosulfan treatment group. The difference between the treatment groups was statistically significant (p=0.0261, adjusted for donor type as factor, and risk group and centre as strata using Cox regression model). The HR was 0.75 (95% CI; 0.59, 0.97) in favour of treosulfan.

B.2.10.5 Deterioration of the KPS to less than 60 points

The deterioration of KPS <60 points was also analysed, since values <60 points are considered to be of major clinical relevance. At the time of analysis, 44 patients (15.5%) in the busulfan treatment group and 23 patients (8.5%) in the treosulfan treatment group deteriorated to <60 points in the KPS. The Kaplan-Meier estimate at 24 months was 18.8% (95% CI; 14.2%, 24.7%) for the busulfan treatment group and 9.9% (95% CI; 6.5%, 14.9%) for the treosulfan treatment group. The difference between the treatment groups was statistically significant (p=0.0130, adjusted for donor type as factor, and risk group and centre as strata using Cox regression model). The HR was 0.52 (95% CI; 0.31, 0.87) in favour of treosulfan.

B.2.10.6 Safety conclusions

Overall, data from the safety of Mc-FludT.14/L Trial II of the final analysis were similar for patients in the 2 treatment groups, and consistent with the data from the confirmatory analysis.[60]

The incidence of total TEAEs (busulfan 96.1%, treosulfan 92.6%) and subcategories of TEAEs, such as related to IMP (busulfan 67.8%, treosulfan 63.0%), was comparable between the 2 treatment groups. The incidence of SAEs (busulfan 7.1%, treosulfan 8.5%) was also similar in the 2 treatment groups.[60]

No unknown risks were identified and the incidence of significant AEs, CTCAE grade III or IV HSOS, seizures, and blood bilirubin increased, was similar in the 2 treatment groups.[60]

A reduced frequency of treatment emergent adverse events was demonstrated for “Mucositis oral” (busulfan 47.7%, treosulfan 37.8%) and “Nausea” (busulfan 41.0%, treosulfan 33.0%). The incidence of Mucositis CTCAE grade III or IV AEs was similar in the 2 treatment groups (busulfan 7.4%, treosulfan 5.9%).[60]

The incidence of acute GvHD was similar in the 2 treatment groups, with a slight numerical advantage in favour of treosulfan (busulfan 57.2%, treosulfan 52.6%), that was not statistically significant. A similar trend was seen for acute GvHD grade III-IV, where the incidence was slightly higher in the busulfan treatment group (busulfan 8.1%, treosulfan 6.3%), but there was no statistically significant difference.[60]

Chronic GvHD incidence was similar in the 2 treatment groups (busulfan 59.5%, treosulfan 60.3%). There was a numerical advantage in favour of treosulfan for extensive chronic GvHD (busulfan 26.7%, treosulfan 19.7%).[60]

Overall, the incidence of GvHD-related deaths was not statistically different, with a slight numerical advantage in favour of treosulfan (busulfan 7.4%, treosulfan 4.8%).[60]

Laboratory parameters were comparable for the 2 treatment groups throughout the study.

Vital signs were comparable for the 2 treatment groups throughout the study. The median.

KPS following transplantation was slightly in favour of treosulfan. In addition, clinically relevant exploratory endpoints “time to deterioration of KPS by at least 20 points” as well as “deterioration of the KPS to less than 60 points” were significantly in favour of the treosulfan treatment group.[60]

B.2.11 Ongoing studies

As of May 2019, one medac sponsored study in paediatric patients with non-malignant disease is currently ongoing (MC-FludT.16/NM).[68] For this indication in this age group, the CHMP has deferred its opinion until mature data are available. The final results of this study are anticipated in December 2019.

B.2.12 Innovation

Although treosulfan is not a new chemical entity, its impact on the conditioning treatment for alloHSCT in patients with malignant disease is profound. Treosulfan has a well-established safety profile as it has been available as palliative treatment for epithelial ovarian cancer for over 30 years. Treosulfan for infusion was first authorised for the treatment of ovarian cancer in 1990. Since 1996 (when electronic storage of post-marketing exposure data was launched at medac), approximately 21,000 patients have been treated with this agent.

However, treosulfan’s innovative potential lies in offering a reduced treatment-related toxicity agent for transplant conditioning that allows effective engraftment, has reduced the non-relapse mortality and has translated into better survival rates.

In fact, data from the pivotal phase III, prospective, randomised study, MC-FludT.14/L Trial II demonstrated that treosulfan was statistically significantly non-inferior to busulfan with regard to event-free survival (EFS) at 24 months after alloHSCT (65.7% vs 51.2%, p=0.0000001), with a statistically significant survival difference in favour of treosulfan (p=0.001). Also, in comparison to busulfan, treosulfan statistically significantly improved, complete donor type chimerism at day +28 (93.2% vs 83.3%; p=0.0159), OS (72.7% vs 60.2%; p=0.0037), NRM (12.0% vs

20.4%; p=0.0343), TRM (12.8% vs 24.1%; p=0.0043), GvHD-free and relapse/progression-free survival (GRFS) (50.3% vs 37.1%; p=0.0087) and chronic GvHD and relapse/progression-free survival (CRFS) (51.4% vs 37.2%; p=0.0030) at 24 months after alloHSCT.[60]

A treosulfan-based myeloablative RTC regimen avoids the toxicity of conditioning therapy in patients unable to tolerate traditional MAC, without sacrificing efficacy in terms of engraftment or relapse rate and ultimately leads to a clinically meaningful overall survival benefit.

B.2.13 Interpretation of clinical effectiveness and safety evidence

The safety and efficacy data from on the large pivotal Phase III study, MC-FludT.14/L Trial II,[60] as presented in this submission, are robust and demonstrate that conditioning therapy with treosulfan in combination with fludarabine can be successfully used for adult patients with AML and MDS who are considered ineligible for standard MAC. Specifically, the treosulfan/fludarabine regimen was demonstrated to be statistically significantly non-inferior to the busulfan/fludarabine RIC regimen with regard to EFS at 24 months after alloHSCT (p=0.0000001). The treatment difference for EFS was statistically significantly in favour of treosulfan (p-value for superiority vs busulfan=0.0005787), indicating a clinically relevant long-term advantage of treosulfan over busulfan.[60]

Results from the primary endpoint analysis based on 570 patients[60] are consistent with those from the interim analysis based on 476 patients.[72] The statistically significant advantage of treosulfan compared with busulfan was also demonstrated for the majority of the secondary endpoints: complete donor type chimerism (p=0.0381), OS (p=0.0037), NRM (p=0.0343), TRM (p=0.0043), GRFS (p=0.0087), and CRFS (p=0.0030). However, the cumulative incidence of relapse/progression at 24 months was comparable between the two regimens.[60]

In addition, a clinically relevant positive effect of treosulfan on patient functional ability, status of general well-being and activities of daily life was shown with treosulfan. Post-surveillance data demonstrated the clinically relevant long-term benefit of treosulfan up to 4 years after HSCT.[60]

Safety analysis revealed that no unknown risks were identified and overall frequencies of AEs and SAEs were comparable to busulfan.[60]

The pivotal trial compared treosulfan with busulfan. Whilst other conditioning regimens are used, busulfan-based conditioning is the most frequently used and widely accepted reduced intensity conditioning (RIC) regimen especially for patients with AML and MDS.[15–17] Importantly, the choice of the busulfan comparator for this study was confirmed by the European Medicines Agency (EMA) during scientific advice procedures with medac. Preliminary registry data (n=45 patients; 21 treosulfan and 24 melphalan patients) which has been shared with medac (Appendix L) by Dr Patel shows the feasibility of treosulfan-based conditioning and that survival with treosulfan conditioning is numerically improved compared to melphalan-based conditioning.

Whilst the pivotal trial results demonstrate treosulfan's safety and efficacy in patients with AML and MDS, it is thought that treosulfan conditioning should not be reserved to treat defined malignancies within certain age groups. Key opinion leaders (See Appendix L) consider that the existing trial data is sufficiently compelling to support the use of treosulfan conditioning in all patients with malignant diseases requiring treatment with myeloablative conditioning prior to alloHSCT.

With regard to the age groups of patients, in addition to the pivotal study and the three supportive phase II studies in adults, there is also phase II data from the MC-FludT.17/M trial[63] in children with malignant haematological diseases which demonstrated low treatment-related mortality, sustained engraftment and chimerism, low relapse rate, and favourable OS after alloHSCT.

In summary, in adults with AML or MDS, the combination treosulfan/fludarabine when compared to a RIC busulfan/fludarabine regimen has been shown to:[59,60,72]

- significantly reduce NRM and TRM
- significantly improve survival outcomes (i.e. EFS, OS and GvHD-free survival.

B.3 Cost effectiveness

Summary of cost-effectiveness results

De novo cost-effectiveness model

- NICE has not previously appraised HSCT conditioning therapies in AML and MDS and therefore a de novo economic model was developed utilising the same patient population as in the clinical study report from the treosulfan pivotal trial (MC-FludT.14/L Trial II) which is believed to be reflective of the UK population.[60]
- A partitioned-survival model structure was considered the most appropriate and in line with two recent prior AML technology appraisals (TA523[91] & TA552[92]).
- Alternative modelling options were considered but rejected:
  - A decision tree model was considered to be too simplistic a structure to accurately capture the long-term benefits of a conditioning therapy that impacts the effectiveness of the whole HSCT procedure.
  - A semi-Markov approach was critiqued in two recent AML technology appraisals as being unnecessary complex and the Evidence Review Group (ERG) suggested a partitioned survival approach.
- Patients enter the model having received an HSCT, and health state occupancy is determined by event-free survival (EFS) and overall survival (OS) curves generated by fitting models to the survival data from the Phase III trial. In this partitioned-survival model, patients can transition between a post-HSCT recovery/remission state (EFS state), a relapsed / progressed disease state, and a death state.
- A cycle length of 28 days matched closely with the estimated average duration of an adult allogeneic HSCT (alloHSCT) procedure inpatient stay and was utilised along with half-cycle correction to minimise potential bias introduced by the choice of cycle length.
- A HSCT is a potentially curative treatment, a lifetime (40 year) time horizon was assumed.
- Costs and outcomes were discounted by 3.5%, as per the NICE reference case.
- The model takes the perspective of the National Health Service (NHS) and Personal Social Services (PSS).

Base case cost-effectiveness results

- In the base case analysis of cost-effectiveness, treosulfan was dominant over busulfan, with lower total costs (treosulfan £137,062 versus busulfan £160,821) and higher total quality-adjusted life-years (QALYs) (treosulfan 6.44 vs 5.55) and an incremental QALY of -0.89.

Sensitivity analyses

- The mean probabilistic incremental total costs were -£19,084 for treosulfan compared to busulfan, with mean probabilistic incremental total QALYs of 0.79 based on 5,000 simulations. The mean incremental total life years were 1.08 for treosulfan compared to busulfan. The mean probabilistic incremental cost-effectiveness ratio (ICER) was -£24,188 (95% CI: -£21,481 to -£16,688 for incremental costs, 0.76-0.82 for incremental QALYs), with a mean incremental net monetary benefit (INMB) of £42,824 (based on a £30,000 per QALY threshold).
- Across all wiliness to pay (WTP) thresholds examined, treosulfan was highly cost-effective, with an 89.9% probability of treosulfan being cost-effective at a WTP threshold of £30,000 per QALY.
- Deterministic sensitivity analysis using a 20% parameter variant illustrated that treosulfan was dominant over busulfan with the exception of the mean log coefficients for the treosulfan OS and busulfan EFS curves, which produced ICERs of £34,821 and £33,040 respectively.
- The majority of scenario analyses showed only relatively minor changes in total incremental costs and QALYs.

Conclusions

- Based on the de novo economic model, treosulfan was highly cost-effective as a conditioning therapy prior to HSCT in patients with AML and MDS.
- The results of this economic evaluation are relevant to the patients studied in the pivotal phase III study, but KOLs have opined (Appendix L) (that the results may also be relevant to patients with other malignancies which were not covered by our model.

B.3.1 Published cost-effectiveness studies

Summary of published cost-effectiveness studies

- The systematic literature review (SLR) found only one cost-effectiveness analysis of reduced-intensity conditioning, Le Corroller et al[93] (see Appendix D).
- A targeted literature search found a further eight studies (Batty et al,[94] Goss et al,[95] Levy et al,[96] Pan et al,[97] Stein et a,l[98] Tremblay et al,[99] and Uyl-de Groot[100]).

A systematic literature review (SLR) was conducted to identify any relevant economic evaluations for the treatment of adult and children with malignant disease

who require conditioning therapy as part of allogeneic haematopoietic stem cell transplantation (alloHSCT). Searches were performed in February 2019 and full details of the SLR search strategy, study selection process, results and quality assessment of included studies are reported in Appendix D.

The systematic literature review identified a cost-effectiveness analysis of two reduced intensity conditioning (RIC) regimens; this was the sole cost-effectiveness analysis relating to haematopoietic stem cell transplantation (HSCT) conditioning therapy.[93] Due to a lack of HSCT conditioning regimen cost-utility studies, economic analyses in acute myeloid leukaemia (AML) and myelodysplastic syndrome (MDS) were searched for in a targeted literature review. A brief summary of the identified economic analyses are presented below in Table 35 on the following page.

Table 35: Summary of cost-effectiveness analyses detected in the systematic review and the targeted literature review

Study	Year	Summary of model	Patient population (average age in years)	QALYs (intervention, comparator)	Costs (currency) (intervention, comparator)	ICER (per QALY gained)
Studies identified in SLR						
Le Corroller et al[93]	2017	Cost-effectiveness analysis of FBA versus FTBI conditioning prior to HSCT in haematological malignancies [Partitioned survival model]	Not reported	Not reported	FBA: €111,725 FTBI: €98,316	FBA versus FBTI: €35,054 per year of PFS gained
Studies identified in targeted literature review						
Batty et al[94]	2014	Cost-effectiveness analysis of decacitabine versus cytarabine and daunorubicin [Semi-Markov model]	60 and older	AD: 0.47 Decitabine: 0.61	AD: $168,863 Decitabine: $108,084	Decitabine versus AD: -$433,756
Crespo et al[101]	2013	Cost-effectiveness analysis of azacitidine for higher-risk	"Advanced age"; not	Azacitidine: 3.06, 3.39, 2.94, 3.11 (versus each	Azacitidine: €107,168, €115,537, €106,422, €108,605	Azacitidine versus BSC: €39,610

Study	Year	Summary of model	Patient population (average age in years)	QALYs (intervention, comparator)	Costs (currency) (intervention, comparator)	ICER (per QALY gained)
		MDS in Spain [Markov model]	explicitly stated	of the below, respectively): BSC: 1.24 LDC: 1.36 SDC: 0.98 ConCR: 1.22	(versus each of the below, respectively): BSC: €35,090 LDC: €53,184 SDC: €59,725 ConCR: €43,170	Azacitidine versus LDC : €30,531 Azacitidine versus SDC : €23,804 Azacitadine versus ConCR : €34,673
Goss et al[95]	2006	Cost-effectiveness of lenalidomide in transfusion-dependent MDS [Structure not specified]	Not reported	Lenalidomide : 0.78 BSC: 0.53	Lenalidomide: $63,385 BSC: $54,940	Lenalidomide versus BSC: $35,050
Levy et al[96]	2014	Cost-effectiveness of azacitadine for higher-risk MDS in Canada [Markov model]	70	Azacitidine: 1.96 ConCRs: 1.06	Azacitidine: $112,354 ConCRs: $35,908	Azacitidine versus BSC: $86,182
Pan et al[97]	2010	Cost-effectiveness of decitabine versus BSC in intermediate / higher-risk MDS in the US [Markov model]	Not explicitly stated; ≥18 years	Decitabine: 0.938 BSC: 0.886	Decitabine: $122,940 BSC: $122,666	Decitabine versus BSC: $5,277
Stein et al[98]	2019	Cost-effectiveness of midostaurin in AML in the USA [Partitioned survival model]	45 years	Midostaurin: 7.30 Placebo: 5.94	Midostaurin: $4,043,470 Placebo: $3,959,741	Midostaurin versus placebo: $61,167
Tremblay et al[99]	2018	Cost-effectiveness of midostaurin versus BSC in AML in the UK	18-59	Midostaurin: 7.79 BSC: 6.32	Midostaurin: £267,325 BSC: £213,253	Midostaurin versus BSC: £36,826

Study	Year	Summary of model	Patient population (average age in years)	QALYs (intervention, comparator)	Costs (currency) (intervention, comparator)	ICER (per QALY gained)
		[Partitioned survival model]				
Uyl-de Groot et al[100]	1998	Cost-effectiveness of GM-CSF + intensive RIC in AML	≥60	GM-CSF: 0.816 BSC: 0.800	GM-CSF: $51,554 BSC: $57,799	No ICER reported

Abbreviations: AML, acute myeloid leukaemia; BSC, best supportive care; ConCR, conventional care regimens; LDC, low dose chemotherapy; MDS, myelodysplastic syndrome; FBA, fludarabine + busulfan + thymoglobulin; FTBI, fludarabine / total body irradiation [2 Grays]; GM-CSF, granulocyte-macrophage colony-stimulating factor; HSCT, haematopoietic stem cell transplant; ICER, incremental cost-effectiveness ratio; QALYs, quality-adjusted life years; RIC, reduced intensity conditioning; SDC, standard dose chemotherapy.

B.3.2 Economic analysis

There have been no previous NICE technology appraisals (TAs) for HSCT conditioning therapies in AML and MDS and therefore no economic models on which to base this submission. Therefore a de novo economic model was produced.

B.3.2.1 Patient population

The patient population considered in the model is the same as in the as in the clinical study report from the treosulfan pivotal trial (MC-FludT.14/L Trial II).[60] The study population is reflective of the UK population covered by the indication. Population parameters are shown below.

Table 36: Population demographics used in the model (pooled AML and MDS patients)

Variable	Busulfan	Treosulfan	Total
N	283	268	551
N (male)	173	162	335
N (female)	110	106	216
Sex (male)	61.13%	60.45%	60.80%
Sex (female)	38.87%	39.55%	39.20%
Age (mean)	59.9	59.3	59.6
Age (SD)	6.0	6.5	6.3
Weight (mean), kg	79.4	80.9	80.2
Weight (SD), kg	17.7	16.7	17.3
n (MRD)	68	62	130
n (MUD)	215	206	421

Variable	Busulfan	Treosulfan	Total
% (MRD)	24.03%	23.13%	23.59%
% (MUD)	75.97%	76.87%	76.41%
BSA (mean), m^2	1.921	1.942	1.931
BSA (SD), m^2	0.241	0.227	0.235
RBC transfusion dependency, n (No)	219	216	435
RBC transfusion dependency, n (Yes)	64	52	116
RBC transfusion dependency, % (No)	77.39%	80.60%	78.95%
RBC transfusion dependency, % (Yes)	22.61%	19.40%	21.05%

Abbreviations: BSA, body surface area; MRD, matched related donor; MUD, matched unrelated donor; RBC, red blood cell; SD, standard deviation.

B.3.2.2 Model structure

B.3.2.2.1 General model structure

The cost-effectiveness analysis was built in Microsoft Excel 2016.

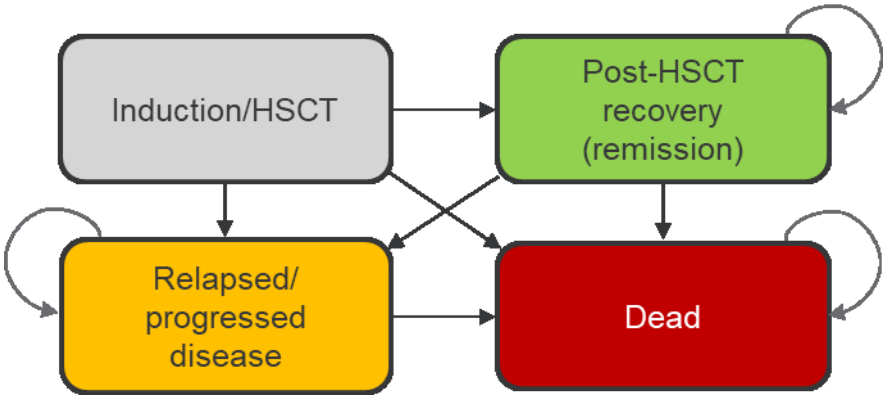
The model structure chosen was a partitioned-survival model, as this was the considered to be most fitting given the clinical data available, and in line with two recent prior AML technology appraisals (TA523[91] & TA552[92]) where HSCT was captured as a later health state in the analysis. As treosulfan is a conditioning therapy that impacts the effectiveness of HSCT, a therapeutic procedure with considerable costs and long-term impacts, a decision-tree model was considered to be too simplistic a structure to accurately capture the long-term benefits of improved conditioning with treosulfan. The Kaplan-Meier plots suggested that assuming exponential event rates would not be appropriate, so a state-transition (Markov) model structure was ruled out.

In addition, while a semi-Markov approach was considered in two recent prior AML technology appraisals (TA399[102] and TA545[103]), the Evidence Review Group (ERG) critiqued the unnecessary complexity of the model used in TA545 and suggested preference for a partitioned survival approach.

Patients enter the model having received an HSCT, and subsequent health state occupancy is determined by event-free survival (EFS) and overall survival (OS)

curves generated by fitting models to the survival data from the Phase III trial (see Section B.3.3.1). In the partitioned-survival model, patients can transition between a post-HSCT recovery/remission state (EFS state), a relapsed / progressed disease state, and a death state. A diagram of the model structure used is shown below in Figure 11.

Figure 11: Model structure diagram



B.3.2.2.2 Cycle length

A cycle length of 28 days was chosen as this matched closely with the estimated average duration of an adult allogeneic HSCT (alloHSCT) procedure inpatient stay based on NHS reference costs[104] (mean duration of 26.8 days) and with the timeframe for which adverse events (excluding acute graft versus host disease (aGvHD) and chronic GvHD (cGvHD)) were reported. This cycle length also matched with cycle lengths used in two recent AML technology appraisals (TA399[102] and TA523[91]), and was considered most suitable for the modelling of treatment costs for relapse patients.

Thirteen 28-day cycles (364 days) were assumed equal to one year for the purposes of the economic analysis. The cycle length of 28 days was used for the entirety of the analysis.

B.3.2.2.3 Half-cycle correction

Despite the short cycle length for the cost-effectiveness analysis, a half-cycle correction was applied to minimise potential bias introduced by the choice of cycle length.

B.3.2.2.4 Time horizon

A lifetime (40 year) time horizon was assumed as HSCT is a potentially curative treatment. Shorter time horizons of 5 and 10 years were considered in scenario analysis.

B.3.2.2.5 Discounting

Discounting of 3.5% per year was applied for both costs and outcomes, as per the NICE reference case.

B.3.2.2.6 Model perspective

The economic analysis was conducted from a National Health Service (NHS) and Personal Social Services (PSS) perspective, as per the NICE reference case.

B.3.2.2.7 Approaches taken in previous technology appraisal submissions

There are no previous TAs for HSCT conditioning therapies in AML and MDS. However, several recent AML appraisals (TA399,[102] TA523,[91] TA545,[103] and TA552[92]) and two MDS appraisals (TA218[105] and TA322[106]) were reviewed as part of the submission.

B.3.3 Clinical parameters and variables

B.3.3.1 Survival modelling

B.3.3.1.1 Introduction to survival modelling

Survival data was recorded for the duration of the clinical trial (up to 1,586 days), but the model requires that survival is estimated over the entire lifetime of the patient population. In order to provide an estimate of survival beyond the end of the clinical trial, models were fitted to the clinical data. The process used for selecting these models is described in detail throughout the rest of this section.

All survival analysis was performed in R using the flexsurvreg, flexsurvcure and flexsurvspline packages, with results validated in Stata 15 (using the streg, strsmix and stpm commands).

B.3.3.1.2 Selection of models for observed survival data

Following review of prior technology appraisals for AML and MDS (e.g. TA545[103]), a variety of modelling approaches were explored to estimate OS and EFS over the duration of the trial. These included:

- Proportional hazards models
- Parametric models
- Mixture-cure models (MCM) / non-mixture-cure models (NMCM)
- Flexible spline models

Proportional hazards could not be assumed following examination of complimentary log-log plots and were deemed inappropriate. Flexible spline models were not included in the economic analysis due to concerns with over-fitting to the observed data. This left standard parametric models and parametric model-derived MCMs / NMCMs to be taken forward.

Survival analyses were carried out for the pooled AML and MDS cohorts as well as the AML and MDS subpopulations separately, stratified by treatment arm. The full survival datasets were used; no other data cuts were considered.

B.3.3.1.3 Base case model selection

Selection of models for survival modelling was informed by NICE Decision Support Unit (DSU) 14.[107] Initially, all parametric models were considered alongside Weibull and lognormal MCMs and NMCMs, and their appropriateness was assessed by the following methods:

- statistically by Akaike information criterion (AIC) and Bayesian information criterion (BIC)
- visual inspection
- Key Opinion Leader (KOL) feedback

The models were primarily selected based on their AIC and BIC performance. The AIC and BIC values for each model for the treosulfan trial arm are shown in Table 37 for OS and Table 38 for EFS. The AIC and BIC values for each model for the busulfan arm are shown in Table 39 for OS and Table 40 for EFS. The models with the lowest AIC/BIC estimates are highlighted in bold.

For both treosulfan and busulfan, the same type of model was used for each arm for consistency and according to advice given in NICE DSU 14.[107] Log-normal non-mixture-cure models were used for all analyses; this model was chosen as it had the lowest AIC for EFS in the pooled AML + MDS population for both arms, as well as the lowest AIC for OS in the pooled AML + MDS population the busulfan arm. The log-normal NMCM was also the second best-fitting model by 0.17 (compared to the Gompertz model) for OS for the AML + MDS patients in the treosulfan arm.

Of the non-cure-based survival models explored, the gamma survival models produced the most consistent estimates. As such, application of the gamma models for both treatments and survival outcomes was explored in scenario analysis.

Table 37: Goodness-of-fit of models in the pooled AML + MDS cohort for treosulfan: OS

Population	AML + MDS		AML		MDS	
Model	**AIC**	**BIC**	**AIC**	**BIC**	**AIC**	**BIC**
Exponential	956.20	959.42	1421.44	1425.04	466.25	468.70
Weibull	952.25	958.69	1412.05	1419.25	462.86	467.75
Lognormal	946.78	953.22	1403.80	**1411.00**	459.94	**464.83**
Loglogistic	950.17	956.61	1409.14	1416.34	461.98	466.87
Gompertz	**946.17**	**952.61**	1404.89	1412.08	462.05	466.94
Gamma	946.90	956.56	**1402.15**	1412.95	**459.77**	467.10
MCM Weibull	948.18	957.83	1407.53	1418.30	463.79	471.08
MCM Log-normal	946.39	956.04	1402.78	1413.56	461.11	468.40
NMCM Weibull	947.82	957.47	1406.88	1417.66	463.49	470.79
NMCM Log-normal	946.34	955.99	1402.74	1413.51	461.09	468.38

Abbreviations: AIC, Akaike information criterion; AML, acute myeloid leukaemia; BIC; Bayesian information criterion; MCM, mixture-cure model; MDS, myelodysplastic syndrome; NMCM, non-mixture-cure model
Bold figures indicate that the model had the lowest AIC/BIC estimate for the associated patient population.

Table 38: Goodness-of-fit of models in the pooled AML + MDS cohort for treosulfan: EFS

Population	AML + MDS		AML		MDS	
Model	**AIC**	**BIC**	**AIC**	**BIC**	**AIC**	**BIC**
Exponential	1141.10	1144.32	1647.18	1650.78	507.98	510.42
Weibull	1119.22	1125.66	1616.80	1623.99	501.50	506.39
Lognormal	1106.67	1113.12	1600.05	1607.34	497.36	502.25
Loglogistic	1114.05	1120.49	1610.12	1617.41	500.05	504.93
Gompertz	1099.75	1106.19	1595.97	1603.26	499.50	504.38

Population	AML + MDS		AML		MDS	
Model	**AIC**	**BIC**	**AIC**	**BIC**	**AIC**	**BIC**
Gamma	1091.17	1100.83	**1580.21**	**1591.15**	**494.80**	**502.13**
MCM Weibull	1103.08	1112.73	1599.83	1610.60	501.29	508.59
MCM Log-normal	1091.38	1101.02	1583.88	1594.65	497.18	504.47
NMCM Weibull	1101.69	1111.34	1597.87	1608.64	500.77	508.06
NMCM Log-normal	**1090.70**	**1100.34**	1582.98	1593.75	497.00	504.29

Abbreviations: AIC, Akaike information criterion; AML, acute myeloid leukaemia; BIC; Bayesian information criterion; MCM, mixture-cure model; MDS, myelodysplastic syndrome; NMCM, non-mixture-cure model
Bold figures indicate that the model had the lowest AIC/BIC estimate for the associated patient population.

Table 39: Goodness-of-fit of models in the pooled AML + MDS cohort for busulfan: OS

Population	AML + MDS		AML		MDS	
Model	**AIC**	**BIC**	**AIC**	**BIC**	**AIC**	**BIC**
Exponential	1039.94	1043.06	1880.66	1884.31	840.45	843.20
Weibull	1039.12	1045.37	1872.45	1879.74	834.41	839.90
Lognormal	1026.82	1033.07	1852.10	1859.39	825.37	830.86
Loglogistic	1033.37	1039.61	1861.97	1869.26	829.36	834.85
Gompertz	1025.49	1031.73	1844.89	1852.18	820.58	**826.07**
Gamma	1016.39	1025.76	1840.41	1851.34	821.70	829.94
MCM Weibull	1017.60	1026.97	1834.22	1845.16	818.80	827.03
MCM Log-normal	1010.96	1020.33	**1831.34**	**1842.27**	818.87	827.11
NMCM Weibull	1016.60	1025.97	1833.20	1844.13	**818.61**	826.85
NMCM Log-normal	**1010.76**	**1020.13**	1831.56	1842.49	819.10	827.34

Abbreviations: AIC, Akaike information criterion; AML, acute myeloid leukaemia; BIC; Bayesian information criterion; MCM, mixture-cure model; MDS, myelodysplastic syndrome; NMCM, non-mixture-cure model
Bold figures indicate that the model had the lowest AIC/BIC estimate for the associated patient population.

Table 40: Goodness-of-fit of models in the pooled AML + MDS cohort for busulfan: EFS

Population	AML + MDS		AML		MDS	
Model	**AIC**	**BIC**	**AIC**	**BIC**	**AIC**	**BIC**
Exponential	1293.49	1296.62	2206.65	2210.29	915.02	917.77
Weibull	1281.64	1287.89	2157.49	2185.62	900.24	905.73
Lognormal	1265.20	1271.45	2150.29	2157.49	888.28	893.77
Loglogistic	1272.44	1278.69	2162.33	2169.62	893.23	898.72
Gompertz	1260.34	1266.59	2137.10	2144.30	879.65	885.14
Gamma	1250.91	1260.28	2125.60	2136.40	878.46	886.70
MCM Weibull	1261.35	1270.72	2136.63	2147.57	879.85	888.09
MCM Log-normal	1250.60	1259.97	2122.69	2133.63	**876.56**	**884.80**

Population	AML + MDS		AML		MDS	
Model	**AIC**	**BIC**	**AIC**	**BIC**	**AIC**	**BIC**
NMCM Weibull	1259.46	1268.83	2134.13	2145.07	879.16	887.39
NMCM Log-normal	**1250.23**	**1259.60**	**2122.51**	**2133.45**	876.69	884.92

Abbreviations: AIC, Akaike information criterion; AML, acute myeloid leukaemia; BIC; Bayesian information criterion; MCM, mixture-cure model; MDS, myelodysplastic syndrome; NMCM, non-mixture-cure model
Bold figures indicate that the model had the lowest AIC/BIC estimate for the associated patient population.

Plots of the survival models versus the original survival data are shown below in Figure 12,

Figure 13, Figure 14, and Figure 15.

Figure 12: AML + MDS pooled population EFS for the treosulfan population

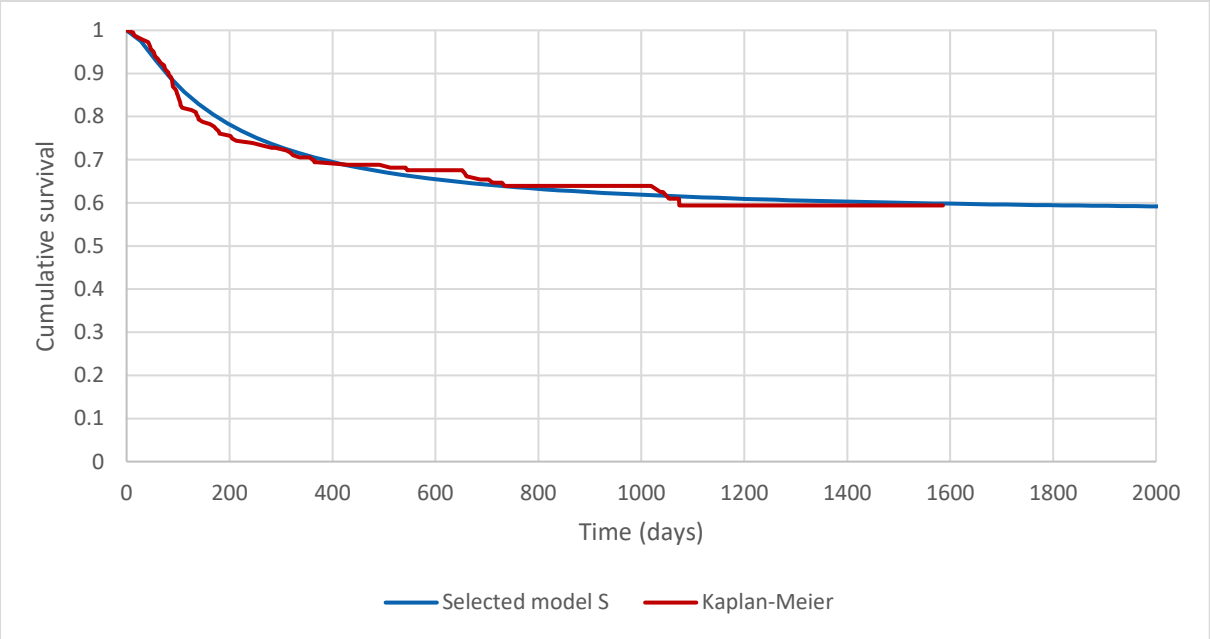


Figure 13: AML + MDS pooled population EFS for the busulfan population

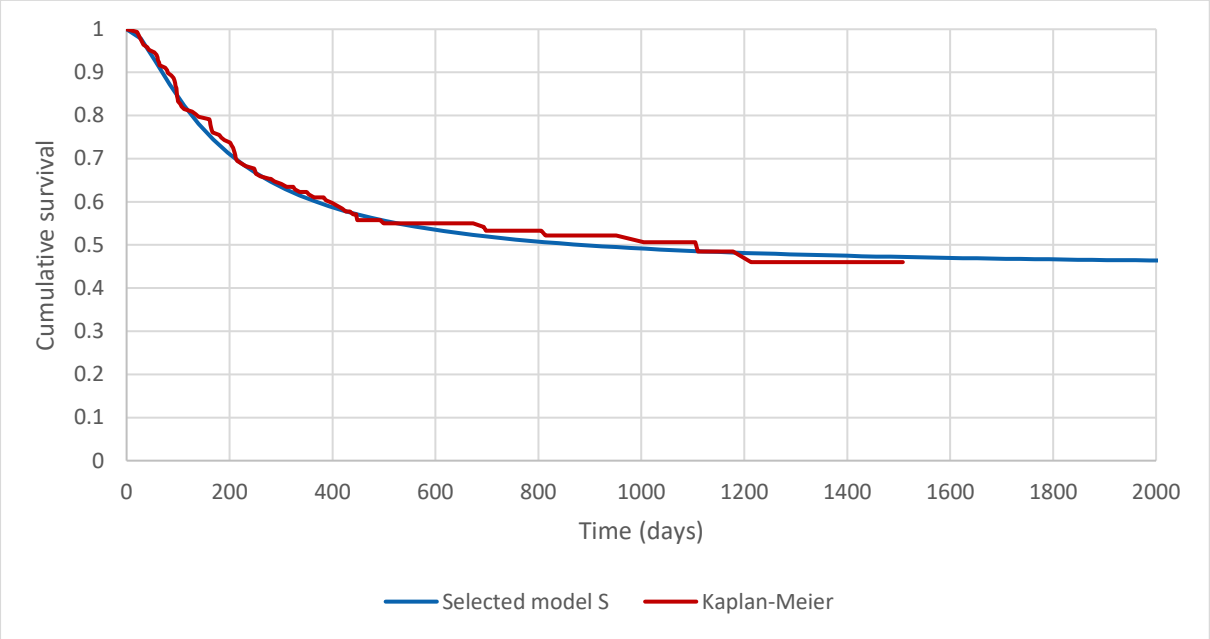


Figure 14: AML + MDS pooled population OS for the treosulfan population

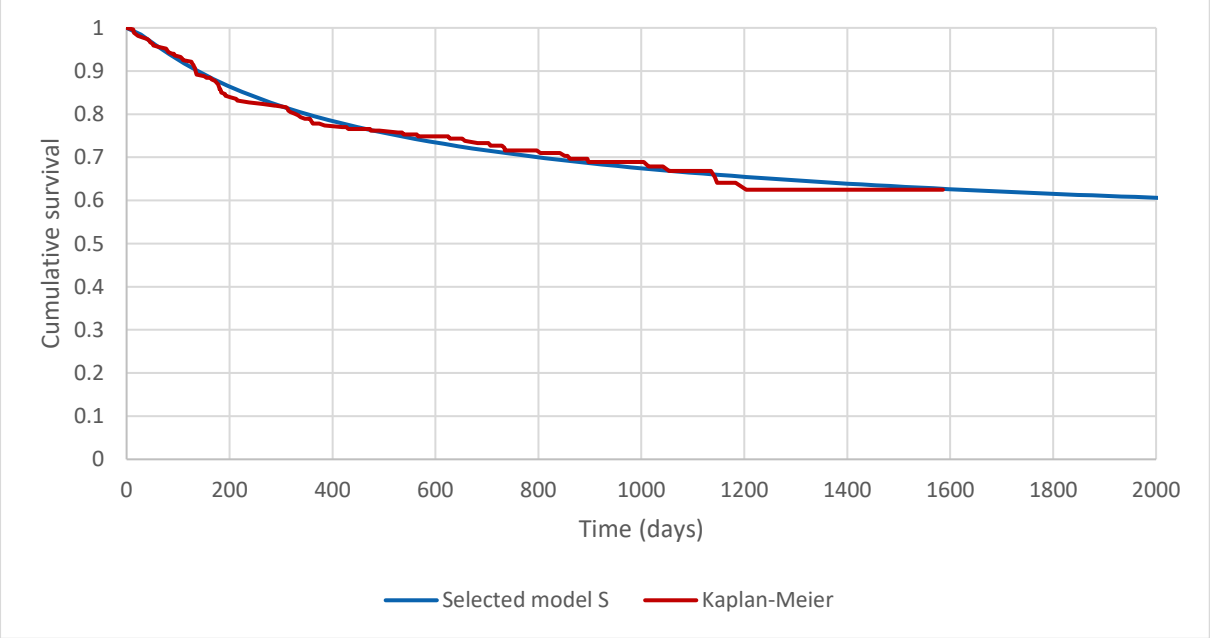
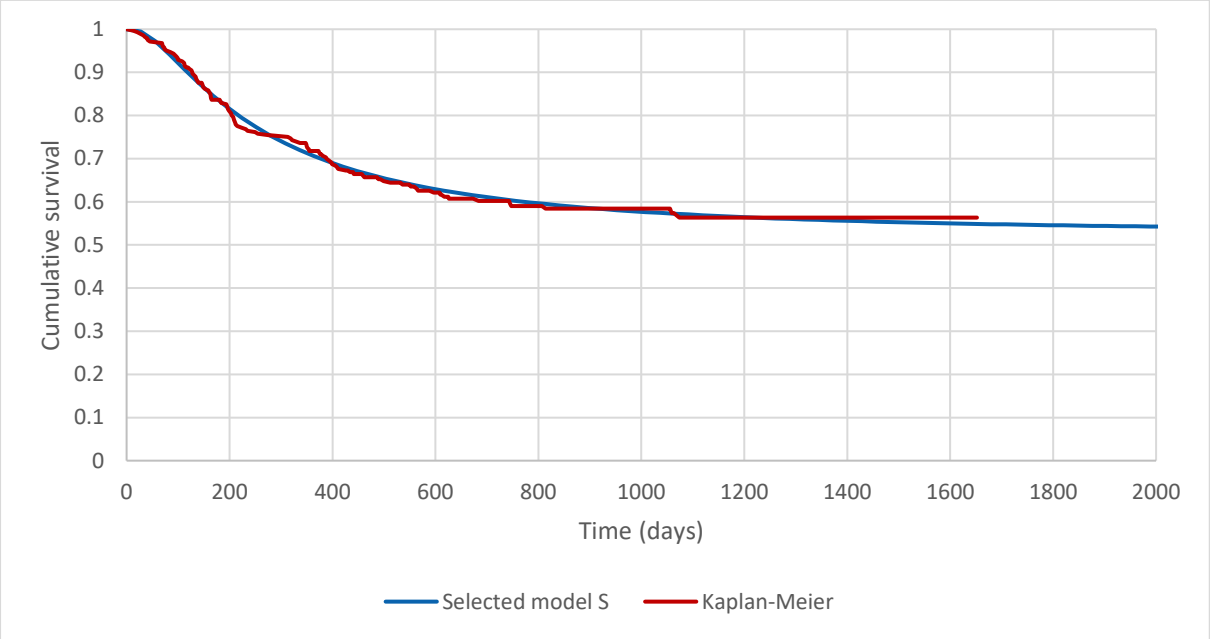


Figure 15: AML + MDS pooled population OS for the busulfan population



B.3.3.1.4 Long-term mortality

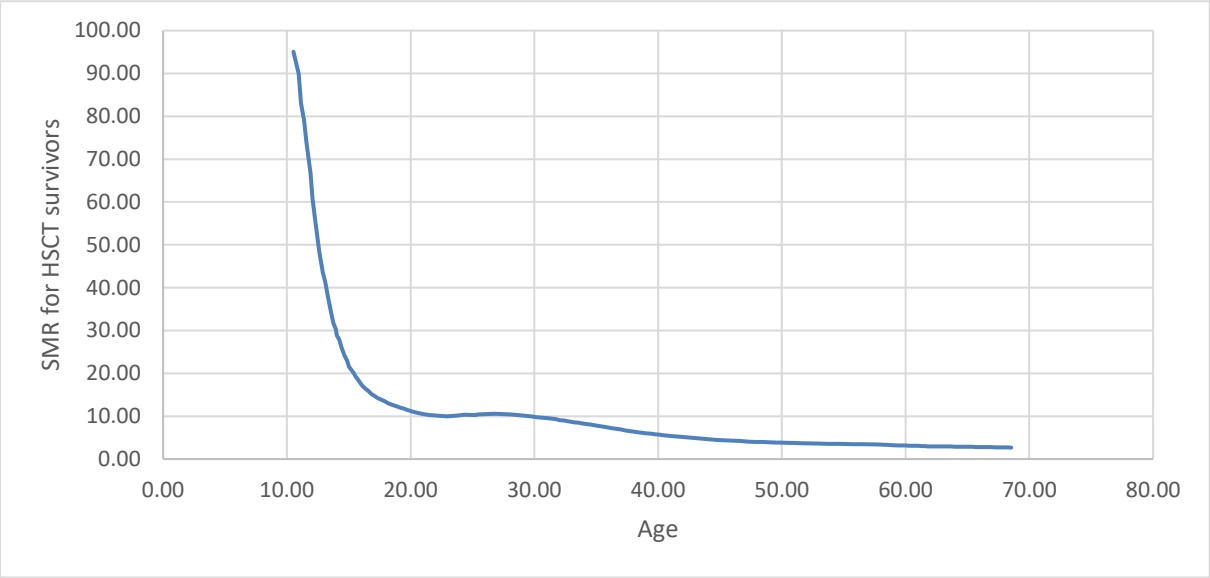
As HSCT is a potentially curative treatment, the option to select a 'cure point' in the model was implemented. Prior to the cure point, the parametric curves for OS and EFS are used. After the cure point, mortality is determined by using life tables for the general population, or life tables with standardised mortality ratios (SMRs) for HSCT applied. In the base case, the cure point was assumed to be 5 years post-HSCT. This was deemed appropriate by a KOL with expertise in the treatment of AML and MDS. This was also corroborated by a second KOL, who added the aetiology of HSCTs is not the main driver of long-term mortality following transplant – complications or long-term outcomes relating to the HSCT itself are. SMRs for HSCT patients were sourced from digitised data for supplementary figure A3C in Martin 2010, which reported SMRs for a cohort of patients who underwent HSCTs for the treatment of ALL, AML, CML, lymphoma, MDS, other haematological malignancies, breast cancer, other unspecified malignancies, aplastic anaemia, and other non-malignant diseases.[108] Several SMRs are included in the model in addition to a HR from TA552;[92] after KOL feedback, the most appropriate mortality ratio for the economic analysis was considered to be 2.3, the HR used in TA552[92] (Table 41).

Table 41: SMRs used to estimate long-term HSCT mortality

Mortality source	Ratio	Comment/description
General population	1.00	Assumes mortality rate equal to UK general population

Mortality source	Ratio	Comment/description
Martin 2010: AlloHSCT	4.30	Overall SMR for alloHSCT patients - from supplementary table A1 in Martin 2010[108]
Martin 2010: Age 45+	3.20	Overall SMR for all patients age 45+ - from supplementary table A3 in Martin 2010[108]
Martin 2010: Mean age + cure point	2.88	Digitised estimate for mean patient age in the model + 5 years - based on supplementary figure A3C in Martin 2010[108]
Martin 2010: TA552 (HR)	2.30	Was considered a plausible estimate by the ERG in the TA552 submission[92]
Martin 2010: Overall HSCT population	4.50	Table 3 in Martin 2010[108]

Figure 16: SMRs for HSCT survivors (digitised from supplementary figure A3C from Martin et al 2010)[109]



If the “mean age + cure point” option is selected, the model utilises the appropriate SMR by adding the cure point onto the mean age of patients in the model cohort, and by selecting the closest SMR from a lookup table based on the above graph (Figure 16); for example, a cure point of 5 years in a cohort with a mean age of 60 years would use the SMR for patients aged 65 (the value for 65.03 in the digitised data, an SMR of 2.88). The author deemed there to be too few data points available to generate meaningful estimates for patients younger than 10 and patients older than 70 years of age based on the spline-smoothed Poisson model that was used. In light of this, it was assumed that patients in the model below the age at the beginning of the curve were assumed to have an SMR equal to the start of the curve (SMR = 95.06), with patients aged above the end of the curve assumed to have an SMR value equal to the end of the curve (SMR = 2.66).

Selection of long-term mortality approach

Five different approaches to long-term mortality were considered. These were:

1. Unadjusted survival as determined by the parametric curves
2. Parametric curves or general population life tables, depending on which has the highest mortality rate
3. Parametric curves or HSCT-specific life tables (as determined by SMRs), depending on which has the highest mortality rate
4. Use of the parametric curves up to a 5-year cure point, followed by switch to general population life table mortality rates
5. Use of the parametric curves up to a 5-year cure point, followed by switch to HSCT-specific life table (as determined by SMRs) mortality rates

Following KOL feedback, approach 5 was deemed to be the most appropriate and reflective of AML and MDS; the KOL approached believed that the capacity for a minority of AML / MDS patients to relapse long after the initial HSCT procedure (>5 years) was important to capture. This approach was used for both the OS and EFS arms in order to allow a small population of late relapsers to exist. This was deemed appropriate by the KOL with the justification that in patients who are 5 years post-transplant, most mortality would be attributed to the HSCT itself rather than relapse of AML / MDS. The OS and EFS curves for the pooled AML+MDS population for treosulfan and busulfan are shown below in Figure 17 and Figure 18.

Figure 17: Base case OS and EFS curves for the pooled AML + MDS population, treosulfan arm

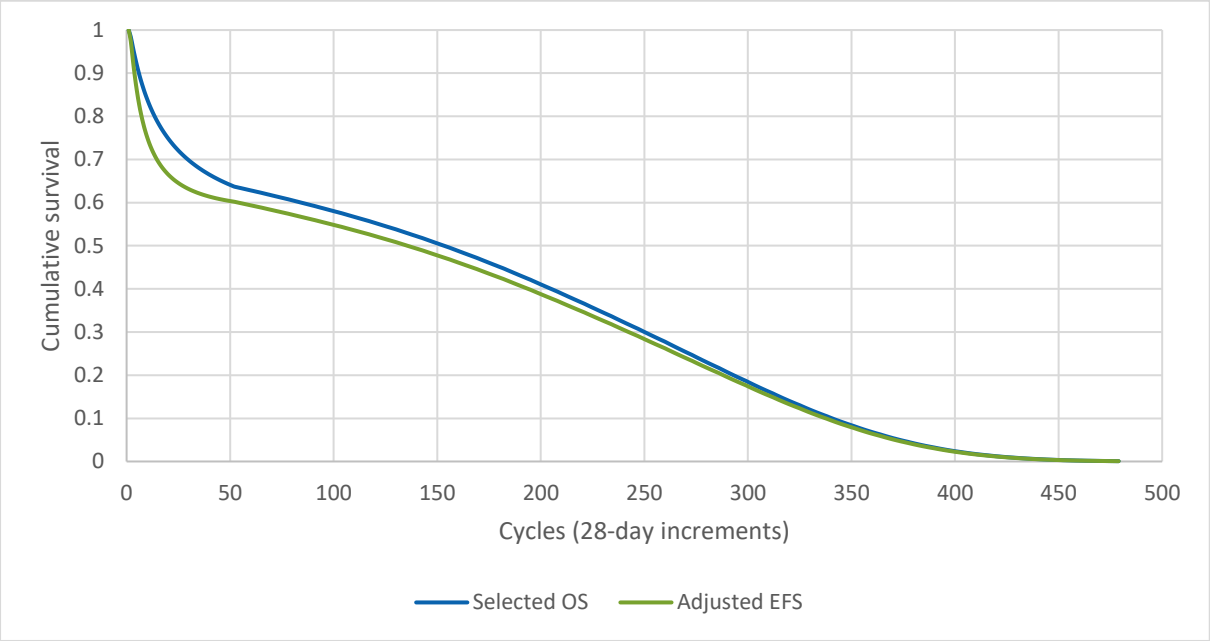
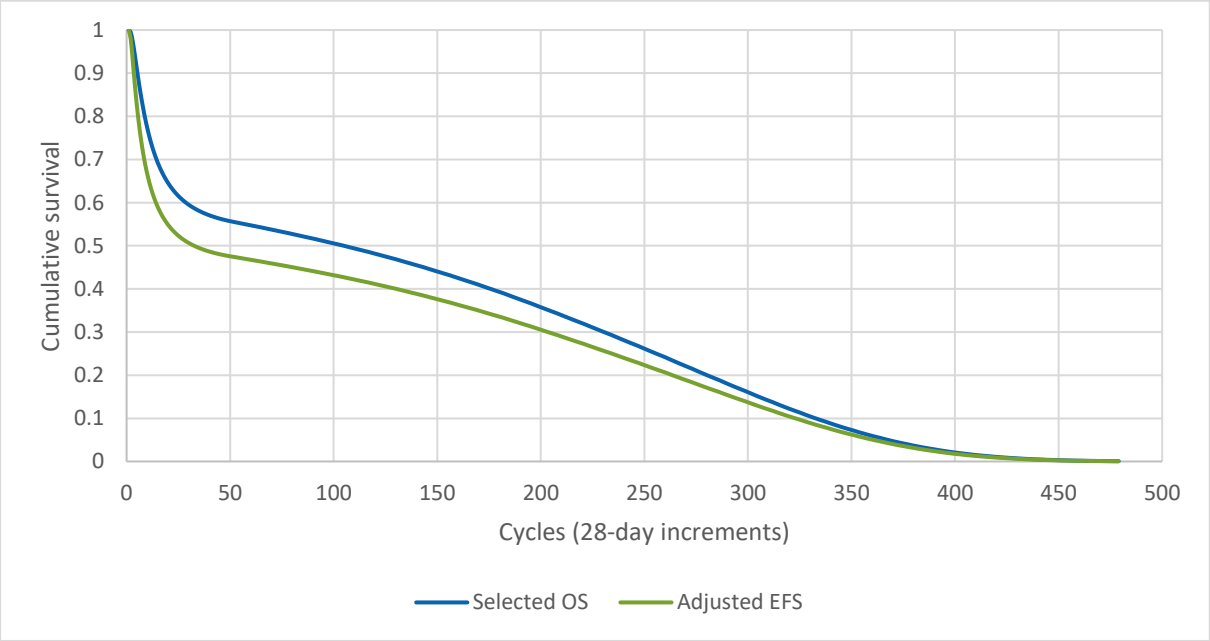


Figure 18: Base case OS and EFS curves for the pooled AML + MDS population, busulfan arm



B.3.3.2 Measurement and valuation of health effects

Health effects were measured in terms of quality-adjusted life-years (QALYs). None of the studies identified during the systematic literature review (SLR), the review of

prior technology appraisals for AML and MDS, and the additional targeted review met the NICE reference case for quality of life data.

Several studies were identified containing European Organization for the Research and Treatment of Cancer Quality of Life Questionnaire-C30 (EORTC QLQ-C30), short form-36 (SF-36) and short form-12 (SF-12) data for AML and MDS patients for the health states described in Section B.3.2.2, which were mapped to EuroQoL 5 Dimensions 3-Level (EQ-5D-3L) using published mapping algorithms. In addition, several preference-based utility estimates were identified that provided custom preference-based utility estimates (using time trade off (TTO) or discrete choice experiment (DCE) methodology), or provided non-UK tariff based EQ-5D-3L estimates.

B.3.3.2.1 Health-related quality-of-life data from clinical trials

Quality of life data were not collected during the phase III trial. As such, published sources of health state utility values (HSUVs) were identified through systematic literature review, as well as review of prior AML and MDS technology appraisals and targeted searching of the literature.

B.3.3.2.2 Health-related quality-of-life studies

As described in Section B.3.3.2.1, quality-of-life data were not collected during the clinical trial. Therefore, published sources of HSUVs were identified through systematic literature review, as well as review of prior AML and MDS technology appraisals and targeted searching of the literature.

Of the two economic modelling papers identified from the systematic literature review, one economic analysis by Kurosawa et al 2011[110] estimating visual analogue scale (VAS) derived health state utility values with suitable health state utility estimates considered for the model. However, only median estimates were provided, and as such this study was not included in the economic analysis.

Fifteen studies on AML, MDS, acute lymphoblastic leukaemia (ALL) and multiple myeloma (MM) patients with preference-based utility measures were identified during the systematic review. Given the patient population for the cost-effectiveness analysis (alloHSCT patients with AML and MDS), studies that exclusively focused on ALL or MM patients (such as Aristides et al 2015[111]) were excluded. Following

exclusion of these studies, one additional abstract and full publication (Stauder et al 2017[112]/Stauder et al 2018[113]) was excluded as the study did not provide fully appropriate utility estimates for the health states in the model. The remaining seven studies (Castejon et al 2018,[114] Joshi et al 2019,[115] Leunis et al 2014,[116] Mamolo et al 2017,[117] Stein et al 2018,[118] Kurosawa et al 2016,[119] Szende et al 2009[120]) were incorporated into the analysis.

In addition, seventeen studies were identified in the SLR that contain data on other quality of life measures. Similar to the preference-based studies, publications that exclusively focused on ALL or MM were excluded from the economic analysis. Of the remaining five studies, one study by Kayastha et al 2018[121] included a measure (Functional Assessment of Cancer Therapy-Leukaemia (FACT-Leu)) that could not be mapped to EQ-5D. Buckley et al 2018[122] was also excluded as the FACT-General (FACT-G) data collected in the study were not explicitly described. Furthermore, Timilshina et al 2016[123] was excluded as the study did not report Quality of Life Questionnaire - Core Questionnaire (QLQ-C30) domain scores and focused on the impact of individual QLQ-C30 domains on mortality. Ramos et al 2017[124] was also excluded as the study focused exclusively on anaemic lower risk MDS patients and did not provide QLQ-C30 or SF-36 domain data for explicit health states relevant to the cost-effectiveness analysis.

The remaining study by Messerer et al 2008, [125] focusing on AML patients in long-term remission was considered for the model. However, this study was subsequently excluded due to concerns around the face validity of both the QLQ-C30 domain scores and mapped EQ-5D estimates for a long-term remission AML patient population, and wide variation in the mapped EQ-5D estimates using the algorithms described in Section B.3.3.2.3 (ranging from 0.42 with McKenzie et al 2009[126] to 0.73 with Proskorovsky et al 2014[127]).

Given the lack of studies meeting the NICE reference case, additional targeted review of the literature and prior AML and MDS NICE technology appraisals was performed. However, no further studies with UK EQ-5D data were identified.

From the targeted review, four additional studies with non-UK tariff EQ-5D data (Kurosawa et al 2015,[128] Kurosawa et al 2014,[129] Slovacek et al 2007,[130] Uyl de

Groot et al 1998[100]), three additional studies with other preference-based utility data (Stein et al 2019,[98] Pan et al 2010,[97] Goss et al 2006[95]), two studies with QLQ-C30 data (Grulke et al 2012,[131] Peric et al 2016[132]), two studies with SF-36 data (Peric et al 2016,[132] Lee 2002[133]) and a single study with SF-12 data (Lee et al 2006[134]) were identified for inclusion in the model.

Furthermore, additional EQ-5D estimates mapped from QLQ-C30 data (using the Proskorovsky et al 2014[127] and McKenzie et al 2009[126] algorithms) were collected from TA399, that were also used for the base case analysis in TA545.[103]

As Pan et al 2010[97] used the TTO utility estimates from Szende et al 2009,[120] this study was subsequently excluded. Slovacek et al 2007[130] was also excluded in the model due to the extremely limited numbers of patients (e.g. 2 patients aged 60-69) used to derive the age-based AML utilities and lack of other utility data for specific health states used in the model.

While two of the studies by Kurosawa et al appeared to be analyses of the same patient population (Kurosawa et al 2016[119], Kurosawa et al 2014[129]), each study was included on an individual basis due to differences in the post-HSCT health state utility estimates reported and estimates of potential disutility for GvHD. Similarly, while Stein et al 2018[118] and Stein et al 2019[98] included DCE based utility estimates based on the same patient population, both studies were incorporated independently in the economic model due to differences in the relapse utilities estimated and inclusion of specific AE disutilities in the Stein et al 2018[118] publication that were not applied in the Stein et al 2019[98] cost-effectiveness study.

B.3.3.2.3 Mapping

Identification of mapping algorithms

Several studies with non-preference-based quality of life data were identified through systematic and targeted literature review. One mapping algorithm for QLQ-C30 was collected from the systematic literature review (Proskorovsky et al 2014[127] only), with the Health Economics Research Centre (HERC) mapping database (Dakin et al 2018[135]), prior technology appraisals in AML/MDS and targeted review of the literature used to source additional mapping algorithms.

Of the HSUV studies identified through systematic and targeted review, several included measures which could not be mapped to EQ-5D due to a lack of published algorithms, such as FACT-Leu and FACT-Bone Marrow Transplant (FACT-BMT). In addition, several studies included FACT-G total score data; while one mapping study was identified with algorithms using FACT-G total score alone (Cheung et al 2009[136]), this mapping study used Singaporean EQ-5D tariffs, and the studies identified through literature review reporting FACT-G quality of life data did not describe explicit health states relevant for the economic analysis. Mapping of FACT-G outcomes to EQ-5D was thus excluded from consideration.

Following exclusion of these studies, several publications with EORTC QLQ-C30, SF-36 and SF-12 data were identified that could be mapped to EQ-5D. As such, algorithms were collected for mapping these outcomes. As only average domain scores were available for the quality of life studies, non-linear models were excluded. In addition, mapping studies were excluded if they did not use UK EQ-5D tariffs in the study, or if the models required patient level data.

QLQ-C30 mapping algorithms

Following application of these exclusion criteria, six QLQ-C30 studies were identified for inclusion in the model. One study (Khan et al 2014[137]) was excluded as intercept coefficients were not provided for the linear mixed model.

The remaining five studies are summarised in Table 42, with coefficients, standard errors, and measures of model performance and predictive performance (those available from more than one study) summarised in Table 43. Due to apparent errors in the specification of the coefficients in the original study, ordinary least squares (OLS) model coefficients for Crott et al 2010[138] were sourced from Tremblay 2018.[99]

Table 42: Summary of QLQ-C30 to EQ-5D mapping study characteristics

Mapping study	Indication/patient group	Country(s) of patient enrolment	Model	N observations in estimation sample	Tests of model performance	Tests of predictive performance
Proskorovsky et al 2014[127]	Multiple myeloma	UK and Germany	OLS	154	Adjusted R^2	RMSE, comparison of predicted and observed utility values overall and by symptom severity group
Crott et al 2010[138]	Breast cancer	International trial population*	OLS	~800	Adjusted R^2	MAE, MAD, RMSE, plots of observed and predicted disutility scores
McKenzie et al 2009[126]	Oesophageal cancer	UK	OLS	877	Adjusted R^2	Absolute differences, RPE, whether observed EQ-5D values lie within 95 % CI of predicted values, minimal clinically important difference (suggested to be 0.03), difference between predicted and observed utilities/QALYs
Kontodimopoulos et al 2009[139]	Gastric cancer	Greece	OLS	48	Adjusted R^2	RMSE, Pearson's correlation coefficients, range of achievable utilities
Marriott et al 2017[140]	Colorectal cancer	International trial population†	Linear mixed effects model	529	Adjusted R^2	MAE, RMSE, plots of fitted values against the observed values as quantile–quantile plots to visually assess fit, testing model fit in each quartile of results

Abbreviations: CI, confidence interval; EQ-5D, EuroQoL 5 dimensions; MAD, median absolute error; MAE, mean absolute error; OLS, ordinary least squares; QALY, quality adjusted life year; RMSE, root mean squared error; RPE, relative prediction error; UK, United Kingdom

*Study notes that centres in five countries (Belgium, France, Netherlands, Switzerland and United Kingdom) were given the option to complete EQ-5D data – however, the patient distribution across countries providing EQ-5D data was not presented

†Patients were recruited from 87 centres in Australia, Europe, Israel, New Zealand and the United States

Table 43: Summary of QLQ-C30 mapping algorithm coefficients, standard errors and measures of model/predictive performance

Domain	Proskorovsky et al 2014[127] (Full model)		Crott et al 2010[138] (Model 3)		McKenzie et al 2009[126]		Kontodimopoulos et al 2009[139]		Marriott et al 2017[140]	
	Coefficient	SE*	Coefficient**	SE†	Coefficient	SE†	Coefficient	SE	Coefficient	SE
Intercept	0.15540	0.12591	0.85928	0.08491	0.23760	0.06988	-0.18143	0.09160	0.39820	0.04920
GH	0.00198	0.00083	-	-	0.00160	0.00052	0.00546	0.00230	0.00080	0.00040
PF	0.00463	0.00116	-0.00697	0.00169	0.00040	0.00040	0.00508	0.00150	0.00190	0.00040
EF	0.00141	0.00077	-0.00873	0.00184	0.00280	0.00045	0.00313	0.00120	0.00140	0.00030
CF	-0.00049	0.00073	-	-	0.00090	0.00041	-	-	0.00100	0.00040
RF	0.00058	0.00077	-	-	0.00220	0.00042	-	-	0.00110	0.00030
SF	0.00060	0.00064	-0.00399	0.00179	0.00020	0.00030	-	-	-0.00030	0.00030
FA	0.00016	0.00091	-	-	-0.00210	0.00056	-	-	-0.00050	0.00040
NV	0.00041	0.00145	-	-	-0.00050	0.00031	-	-	0.00060	0.00040
PA	-0.00249	0.00062	0.00356	0.00049	-0.00240	0.00039	-	-	-0.00210	0.00030
AP	-0.00037	0.00068	-	-	0.00030	0.00026	-	-	-0.00050	0.00030
CO	-0.00050	0.00053	0.00115	0.00033	0.00010	0.00016	-	-	0.00010‡	0.00020
DI	-	-	0.00399	0.00120	-0.00030	0.00048	-	-	0.00010‡	0.00030
DY	0.00060	0.00056	-	-	0.00040	0.00027	-	-	0.00030	0.00030
SL	0.00082	0.00050	-0.00037	0.00080	0.00004	0.00029	-	-	0.00020	0.00020
FI	0.00080	0.00051	-	-	-0.00060	0.00043	-	-	-0.00060	0.00020

Domain	Proskorovsky et al 2014[127] (Full model)		Crott et al 2010[138] (Model 3)		McKenzie et al 2009[126]		Kontodimopoulos et al 2009[139]		Marriott et al 2017[140]	
	Coefficient	SE*	Coefficient**	SE†	Coefficient	SE†	Coefficient	SE	Coefficient	SE
PF squared	-	-	0.00004	0.00001	-	-	-	-	-	-
EF squared	-	-	0.00006	0.00001	-	-	-	-	-	-
SF squared	-	-	0.00003	0.00001	-	-	-	-	-	-
DI squared	-	-	-0.00005	0.00002	-	-	-	-	-	-
SL squared	-	-	0.00001	0.00001	-	-	-	-	-	-
Adjusted R^2	0.7015		0.801		0.611		0.611		0.646	
MAE	-		0.08		-		-		0.127	
RMSE	0.164		0.096		-		0.192		0.092	

Abbreviations: AP, appetite; CF, cognitive functioning; CO, constipation; DI, diarrhoea; DY, dyspnoea; EF, emotional functioning; FA, fatigue; FI, financial difficulties; GH, global health; MAE, mean absolute error; NV, nausea/vomiting; PA, pain; PF, physical functioning; RF, role functioning; RMSE, root mean squared error; SE, standard error; SF, social functioning; SL, sleep/insomnia

*Estimated from p values using a quantile function. P values described as <0.0001 were assumed to be equal to 0.0001

**Sourced from Tremblay et al 2018 due to errors in the original study[99]

†Estimated from t values

‡Coefficients for constipation and diarrhoea were specified as <0.0001 in the study – as such, these coefficients were assumed to be equal to 0.0001

In a linear regression, the probability value (p) is calculated from the t test statistic (t), which is the coefficient β divided by its standard error ($t = \beta/SE_\beta$). The degrees of freedom used in the t-distribution for calculating the p-value are the residual degrees of freedom $SE_\beta = \beta/|t|$. The residual degrees of freedom, on the other hand, are the total degrees of freedom of the variance N−1 minus the model degrees of freedom p−1, where p is the number of parameters including the intercept. The residual degrees of freedom are therefore (N−1) − (p−1) = N−p. From this, we can use the quantile distribution of the t-distribution to calculate the standard error (Hill 1970).[141]

As only p values were available for the coefficients from Proskorovsky et al 2014,[127] standard errors used in the probabilistic sensitivity analysis were estimated from p values using the quantile function of the Student t-distribution, with the analysis performed in R.

The patient population characteristics of the original mapping study (indication, UK patient representation, number of observations in estimation sample) were considered key criteria for selection of the most appropriate mapping algorithm for the model. In addition, the following predictive performance ranking criteria were also considered, based on recommendations from Longworth et al 2014:[142]

1. MAE
2. RMSE
3. MAE on the 0.75 to 1 subset of the EQ-5D-3L
4. MAE on the 0.5 to 0.75 subset of the EQ-5D-3L
5. MAE on the 0 to 0.5 subset of the EQ-5D-3L
6. MAE on the <0 subset of the EQ-5D-3L
7. Error in the mean predicted QALYs compared to the mean observed QALYs

From the algorithms described in Table 42 and Table 43, Proskorovsky et al 2014[127] was selected for the base case analysis. While this study included the second lowest number of estimation sample observations of the five studies and an RMSE value from the original study higher than two other algorithms (Crott et al 2010[138] and Marriott et al 2017[140]), a large proportion of the patient population were from the UK (58%), and the cancer type of the patient population (multiple myeloma) was considered more similar to the model population than the indications for the other

mapping studies, with 12% of the patient population having received a prior autologous stem cell transplant.

In terms of external validation, in a mapping validation study of a large set of mixed cancer patients by Doble et al 2016,[143] the Proskorovsky et al 2014[127] mapping algorithm had a lower mean absolute error than the Crott et al 2010,[138] McKenzie et al 2009[126] and Kontodimopoulos 2009[139] algorithms in the main analysis sample and upper EQ-5D subgroups (0.75 ≤ EQ-5D < 1 and 0.75 ≤ EQ-5D < 1), albeit performing worse than the other algorithms in patients with an EQ-5D score below 0.5 (with the highest MAE for this subgroup). However, since the majority of patients remaining alive in the model over the long term were considered likely to have a utility closer to the general population (i.e. utilities >0.5 up to 100 years of age), this suggested that the Proskorovsky et al 2014[127] algorithm would perform most accurately for the majority of patients and associated health states within the model.

In addition, Proskorovsky et al 2014[127] was judged to have produced the most plausible set of mapped EQ-5D values for the QLQ-C30 study used in the base case analysis (Grulke et al 2012[131]) to estimate post-HSCT recovery/remission utilities, with continuous increases in utilities for each time point post-HSCT and year 4+ post-HSCT utility estimates that were considered to be the most plausible reflection of disutility for functionally cured patients compared to the general population (see Section B.3.3.2.5). Furthermore, Proskorovsky et al 2014[127] was used in the base case analysis for TA399,[102] which was identified as a key source of relapse utilities and adverse event disutilities. As such, application of the Proskorovsky et al 2014[127] algorithm allowed for consistency in application of mapping algorithms used across health states in the model.

From the remaining mapping algorithms, McKenzie et al 2009[126] was selected for scenario analysis due to having the largest total number of observations in the estimation sample (877), an exclusively UK population, the most conservative set of post-HSCT recovery/remission utilities, and internal consistency with use of mapped EQ-5D estimates from TA399.[102] Furthermore, in terms of external validation, McKenzie et al 2009[126] produced a lower MAE (0.144) versus the Crott et al 2010[138] and Kontodimopoulos et al 2009[139] algorithms within a multiple myeloma mapping validation study identified in the systematic literature review (Rowen et al 2012[144]).

SF-36 mapping algorithms

Six studies were identified in the HERC database (Dakin et al 2018[135]) for mapping SF-36 to EQ-5D. One study (Kim et al 2014[145]), was excluded due to the use of Korean EQ-5D tariffs in the study. Another study (Maund et al 2012[146]) was excluded due to concerns around the generalisability of the patient population (frozen shoulder) and reliance on physical and mental component summary scores, which were not available in the SF-36 quality of life studies identified in the literature. Two further studies (Longo et al 2000[147] and Longworth 2007[148]) were also excluded, as the original papers could not be located, and due to concerns around the generalisability of the specific patient populations (breast disease and coronary artery disease respectively) to alloHSCT patients with AML and MDS.

Following exclusion of the studies described above, two available mapping studies were identified with mapping algorithms from SF-36 to EQ-5D for use in the economic analysis. These studies are summarised in Table 44, with coefficients described in Table 45.

Table 44: Summary of SF-36 to EQ-5D mapping study characteristics

Mapping study	Indication/patient group	Country(s) of patient enrolment	Model	N observations in estimation sample	Tests of model performance	Tests of predictive performance
Rowen et al 2009[149]	Hospital inpatients and outpatients	UK	Random effects GLS	33,248	R^2 (overall, between, within)	RMSE
Ara et al 2008[150]	Asthma, chest pain, older people, COPD, irritable bowel syndrome, trauma, back pain, leg disorders, osteoarthritis	UK	OLS	6,350	R^2	MAE, RMSE, ME, number of errors greater than 0.05, minimal important difference in score

Abbreviations: COPD, chronic obstructive pulmonary disease; GLS, generalised least squares; OLS, ordinary least squares; UK, United Kingdom

Table 45: Summary of SF-36 mapping algorithm coefficients, standard errors and measures of model/predictive performance

Domain	Rowen et al 2009 (GLS model 3)[149]		Ara et al 2008 (Model EQ (6))[150]	
	Coefficient	SE*	Coefficient	SE
Intercept	-0.25600	0.05120	-0.169780	0.027100
PF	0.55900	0.11180	0.007710	0.000480
RP	-0.14600	0.02920	-0.000060	0.000080
BP	0.71500	0.14300	0.004880	0.000550
GH	0.40700	0.08140	0.000760	0.000160
VIT	0.01700	0.00340	-0.000340	0.000180
SF	0.29300	0.05860	0.002140	0.000590

Domain	Rowen et al 2009 (GLS model 3)[149]		Ara et al 2008 (Model EQ (6))[150]	
	Coefficient	SE*	Coefficient	SE
RE	0.06700	0.01340	0.000230	0.000090
MH	0.48300	0.09660	0.006090	0.000770
PF squared	-0.22700	0.04540	-0.000040	0.000000
RP squared	0.00100	0.00020	-	-
BP squared	-0.33000	0.06600	-0.000020	0.000000
GH squared	0.03200	0.00640	-	-
VIT squared	-0.01200	0.00240	-	-
SF squared	-0.16300	0.03260	-0.000010	0.000000
RE squared	0.03400	0.00680	-	-
MH squared	-0.24200	0.04840	-0.000030	0.000010
PF×RP	0.02200	0.00440	-	-
PF×BP	-0.03200	0.00640	-	-
PF×GH	0.07300	0.01460	-	-
PF×VIT	-0.13200	0.02640	-	-
PF×SF	-0.02300	0.00460	-	-
PF×RE	0.04700	0.00940	-	-
PF×MH	-0.01400	0.00280	-	-
RP×BP	0.01900	0.00380	-	-

Domain	Rowen et al 2009 (GLS model 3)[149]		Ara et al 2008 (Model EQ (6))[150]	
	Coefficient	SE*	Coefficient	SE
RP×GH	0.06800	0.01360	-	-
RP×VIT	0.05000	0.01000	-	-
RP×SF	0.06700	0.01340	-	-
RP×RE	-0.01200	0.00240	-	-
RP×MH	0.02200	0.00440	-	-
BP×GH	-0.21700	0.04340	-	-
BP×VIT	-0.00200	0.00040	-	-
BP×SF	0.05500	0.01100	-	-
BP×RE	-0.03800	0.00760	-	-
BP×MH	0.13100	0.02620	-	-
GH×VIT	-0.06600	0.01320	-	-
GH×SF	-0.15700	0.03140	-	-
GH×RE	-0.03300	0.00660	-	-
GH×MH	-0.08400	0.01680	-	-
VIT×SF	0.14300	0.02860	-	-
VIT×RE	-0.02000	0.00400	-	-
VIT×MH	0.02300	0.00460	-	-
SF×RE	-0.02300	0.00460	-	-

Domain	Rowen et al 2009 (GLS model 3)[149]		Ara et al 2008 (Model EQ (6))[150]	
	Coefficient	SE*	Coefficient	SE
SF×MH	-0.06500	0.01300	-	-
RE×MH	-0.04800	0.00960	-	-
Age	-	-	-0.001000	0.000200
Age squared	-	-	0.000001	0.000000
R^2	0.71		0.5868	
MAE	-		0.1299	
RMSE	0.15		0.1780	

Abbreviations: BP, bodily pain; GH, global health; MAE, mean absolute error; MH, mental health; PF, physical functioning; RE, role emotional; RMSE, root mean squared error; RP, role physical; SE, standard error; SF, social functioning; VIT, vitality

*Standard errors were estimated as 20% of the coefficient, due to a lack of precision data described in the original study

For the two cGvHD quality of life studies with SF-36 data considered in the economic analysis (Peric et al 2016[132] and Lee et al 2002[133]), both mapping algorithms produced plausible EQ-5D values compared to the general population EQ-5D for the average age of the original populations.

While the SF-36 mapping algorithms were not applied in the base case analysis, the Rowen 2009[149] algorithm was considered more suitable for the analysis due to the larger observation estimation sample size, lower RMSE and more appropriate patient population (hospital inpatients and outpatients) for alloHSCT patients with AML and MDS.

SF-12 mapping algorithms

Eight studies were identified in the HERC database (Dakin 2018[135]) for mapping SF-12 to EQ-5D. The single study with SF-12 data (Lee et al 2006[134]) considered in the model, which studied the quality of life of cGvHD and aGvHD/cGvHD overlap patients, only provided average physical and mental composite scores, and as such this was used as a primary criteria for selecting the most appropriate mapping algorithm.

Following application of this criteria, and exclusion of mapping studies with insufficient description of the mapping algorithms (Rivero-Arias et al 2010),[151] using non-UK EQ-5D tariffs, with non-linear models, or those relying on patient level data or domain scores, three SF-12 mapping studies were identified. One of the three studies, Franks et al 2003,[152] was excluded as the study focused solely on low income individuals.

The remaining two studies are summarised in Table 46, with coefficients described in Table 47. Lawrence et al 2004,[153] did not provide standard error or other precision estimates for the six-variable model; therefore standard errors were estimated as 20% of the coefficient value.

Table 46: Summary of SF-12 to EQ-5D mapping study characteristics

Mapping study	Indication/patient group	Country(s) of patient enrolment	Model	N observations in estimation sample	Tests of model performance	Tests of predictive performance
Franks et al 2004[152]	General population	US	OLS	12,998	R^2	Comparison of observed and predicted EQ-5D values
Lawrence 2004[153]	General population	US	OLS	7,313	R^2	Mallow's C_p, comparison of observed and predicted EQ-5D values

Abbreviations: OLS, ordinary least squares

Table 47: Summary of SF-12 mapping algorithm coefficients, standard errors and measures of model/predictive performance

Domain	Franks et al 2004[154] (SF-12 items only model)		Lawrence et al 2004[153] (6-variable model)	
	Coefficient	SE	Coefficient	SE*
Intercept	0.84690	0.00294	-1.69840	0.33968
PCS	0.01261	0.00024	0.07927	0.01585
MCS	0.00759	0.00025	0.02859	0.00572
PCS squared	-0.00009	0.00002	-0.00141	0.00028
MCS squared	-0.00015	0.00002	-0.00014	0.00003
PCS×MCS	-0.00015	0.00002	-0.00013	0.00003
PCS cubed	-	-	0.00001	0.00000

Domain	Franks et al 2004[154] (SF-12 items only model)		Lawrence et al 2004[153] (6-variable model)	
	Coefficient	SE	Coefficient	SE*
R^2	0.626		0.628	
MAE	-		-	
RMSE	-		-	

Abbreviations: MAE, mean absolute error; MSC, mental composite score; RMSE, root mean squared error; PCS, physical composite score; SE, standard error
*Standard errors were estimated as 20% of the coefficient, due to a lack of precision data described in the original study

Although the studies were developed from a large US general population (US Medical Expenditure Panel Survey), both utilised UK EQ-5D tariffs for the development of the mapping algorithms. Both studies provided limited data on predictive performance, and both models had similar R^2 statistics. While two studies utilising both SF-12 algorithms were identified (Pickard et al 2005[155] and Snedecor 2009[156]), neither study provided a formal comparison of predictive performance of the two algorithms, or applied the algorithms in a patient population that was considered sufficiently similar to alloHSCT patients with AML and MDS.

Mapped SF-12 estimates were not considered in the base case analysis – however, the Franks et al 2004[154] algorithm was considered slightly more appropriate over Lawrence et al 2004[153] due to a larger number of estimation sample observations and provision of standard error estimates for the coefficients. Both mapping studies were included in order to allow for potential exploration of disutility specific to aGvHD (applied as a disutility vs cGvHD).

B.3.3.2.4 Adverse reactions

Although no studies meeting the NICE reference case were identified, several studies were collected from the literature review with utility estimates or quality of life data for GvHD (Castejon et al 2018,[114] Kurosawa et al 2016,[119] Kurosawa et al 2015,[128] Kurosawa et al 2014,[129] Peric et al 2016,[132] Lee et al 2002[133], Lee et al 2006[134]). No study identified, however, included health state utility estimates or quality of life data specific to extensive chronic GvHD (cGvHD) or stage III/IV acute GvHD (aGvHD), although two studies were identified (Peric et al 2016,[132] Lee et al 2002[133]) with SF-36 data using other measures of cGvHD severity.

In Peric et al 2016,[132] definitions of “inactive” and “active/highly active” were used to differentiate patients within the cGvHD group. However, given that clinical experts were unfamiliar with these activity based cGvHD classifications and unsure of the overlap in definition with extensive cGvHD (Shulman et al 1980[157]), mapped estimates were not utilised in the base case or scenario analyses.

For the mild/moderate/severe cGvHD classification (Filipovich et al 2005[158]) used in Lee et al 2002,[133] clinical expert input suggested that patients with moderate to severe cGvHD would likely be similar to those with extensive cGvHD, and that

application of a moderate to severe cGvHD disutility for extensive cGvHD patients would be a reasonable assumption. However, estimates from Lee et al 2002[133] were not used in the base case or scenario analyses due to the lack of a “no cGvHD” reference group for estimating disutility and concerns relating to appropriately weighting the moderate and severe patient group estimates to produce an overall moderate to severe cGvHD utility estimate.

Furthermore, only one study identified (Lee et al 2006[134]) contained quality of life data for patients with aGvHD, and only included data specific to patients with overlapping symptoms of aGvHD and cGvHD. Given concerns regarding application of quality of life data for patients with overlapping aGvHD and cGvHD to patients with only aGvHD, and internal consistency of the mapped EQ-5D estimates of patients with cGvHD only compared to patients with no cGvHD at six months post-HSCT, specific estimates for aGvHD from Lee et al 2006[134] were not included in the base case or scenario analyses. As such, the same disutility was applied for extensive cGvHD and stage III/IV acute GvHD in the base case model, which was considered a reasonable assumption by clinical experts.

Three studies (Peric et al 2016,[132] Lee et al 2002[133] and Lee et al 2006[134]) required mapping from other quality of life measures to EQ-5D. Two studies (Peric et al 2016,[132] Lee et al 2002[133]) contained SF-36 domain score data which were mapped to EQ-5D using algorithms described in Section B.3.3.2.3, Mapping (Rowen et al 2009[149], Ara et al 2008[150]). The third study (Lee et al 2006[134]) was mapped from SF-12 physical and mental composite summary scores using Franks et al 2004[154] and Lawrence et al 2004.[153] While Peric et al 2016[132] also contained QLQ-C30 data, the diarrhoea symptom scores appeared to be missing for patients without cGvHD and the overall cGvHD patient group in the study, and the mapped estimate using Crott et al 2010[138] (0.156) was inconsistent with that described in TA523 (0.173).[91]

For the base case analysis, a GvHD disutility of 0.120 was applied based on Kurosawa et al 2016,[119] which was considered most closely aligned to the NICE reference case and more appropriate than applying mapped EQ-5D estimates from SF-36 and SF-12 using algorithms developed in non-cancer specific populations. A GvHD disutility of 0.190 from Castejon et al 2018[114] was applied in the TTO/DCE utility scenario analysis.

In terms of the remaining adverse events included in the model, published studies on the impact of Grade 3+ CTCAEs on quality of life for AML and MDS patients receiving alloHSCT were extremely limited. Only one of the published studies identified through systematic literature review (Stein et al 2018[118]) produced AE disutility estimates using DCE methodology for adverse events relevant to the model (serious infections, severe diarrhoea, severe redness/skin peeling and abnormally low blood counts). However, two sets of disutilities for Grade 3 and 4 adverse events (AEs) were identified from TA399[102] (0.0240 and 0.0207) based on different mappings of the QLQ-C30 data from the clinical trial, which were judged to be more aligned with the NICE reference case, despite concerns regarding potential differences in the AEs experienced by patients on a chemotherapy treatment (azacitidine) versus patients receiving HSCT. Therefore, these disutilities were applied in the base case model for all other adverse events, with values from Stein et al 2018[118] explored in scenario analysis.

Disutilities for each adverse event were spread over the mean duration of the event and were modelled independently, with the negative QALYs subtracted from the total QALY estimates calculated for the model health states. Duration and incidence estimates are described in Section B.3.4.2.

B.3.3.2.5 Health-related quality-of-life data used in the cost-effectiveness analysis

All patients in the model start in the induction/HSCT health state, with the same HSCT induction health state utility applied to all patients in the first cycle. For patients remaining event-free in the HSCT-recovery (remission) health state after the first cycle, time-based utilities were included to more accurately reflect continuous/stepwise improvement in quality of life following the HSCT procedure. Discharge utilities were applied for a single cycle post-HSCT, followed by individual utility estimates for short-term HSCT recovery (≤6 months and 7-12 months) and long-term HSCT recovery (years 2, 3 and 4+).

For patients experiencing a relapse after the first cycle, utilities for AML and MDS based relapse were individually applied. While separate utilities for AML and MDS relapses respectively were allowed in the model, the same relapse/progression utility was applied to both populations in the base case analysis due to limitations with data

for MDS relapse/progression patients, and based on clinical expert feedback that AML and MDS relapse/progression patients would be likely to have a similar health state utility. However, a TTO-based MDS specific relapse utility estimate (Szende et al 2009[120]) was applied in a scenario analysis, generated through a utility multiplier applied vs short-term HSCT recovery (≤6 months). Although this estimate was generated based on a comparison of health state utilities for patients with and without red blood cell (RBC) transfusion dependence, clinical expert feedback suggested that application of a RBC transfusion dependence based utility to MDS relapse/progression patients was a reasonable assumption.

For the base case analysis, Grulke et al 2012[131] was considered the most suitable study for estimating post-HSCT recovery utility, and contained considerably larger sample sizes of patients for most health states than other studies (with approximately 2,800 patients in total). The specification of five time-based health states for patients in post-HSCT was noted as a key strength of the study during discussion with clinical experts, with more detailed time-based progression of health state utility values regarded as more plausible than the application of one or two health state utilities post-HSCT. QLQ-C30 domain scores for each health state from Grulke et al 2012[131] are described in Table 48.

Table 48: Summary of QLQ-C30 domain scores and health state data sample information from Grulke et al 2012[131]

Domain	During hospitalisation		At discharge		Up to 6 months after HSCT		7-12 months after HSCT		1-3 years after HSCT		>3 years after HSCT	
	Mean	SD	Mean	SD	Mean	Mean	Mean	SD	Mean	SD	Mean	SD
GH	38	18	53	19	60	21	67	22	72	21	74	21
PF	54	29	64	21	72	21	81	21	81	20	86	18
EF	66	27	65	23	75	23	74	22	74	22	74	25
CF	68	28	69	24	79	24	82	25	79	21	77	26
RF	34	37	41	36	54	33	75	28	72	30	78	28
SF	52	32	52	31	63	30	76	24	73	25	77	28
FA	70	30	58	24	49	27	33	22	32	25	30	25
NV	63	32	31	30	14	21	8	18	7	13	4	11
PA	47	38	34	29	21	25	21	24	22	26	16	23
AP	76	34	53	34	33	32	11	16	10	18	8	16
CO	19	30	11	23	6	15	7	16	9	18	8	19
DI	49	39	38	33	16	24	12	19	12	20	10	18
DY	24	35	29	32	27	29	26	32	19	19	20	24
SL	50	38	44	31	25	29	23	30	26	28	26	30
FI	35	39	31	33	36	34	23	31	25	28	23	32
Number of patients	758-1090	355-719	49	49	93-127	93-127	134-225	134-225	551-753	459-661*	692-734	355

Domain	During hospitalisation		At discharge		Up to 6 months after HSCT		7-12 months after HSCT		1-3 years after HSCT		>3 years after HSCT	
	Mean	SD	Mean	SD	Mean	Mean	Mean	SD	Mean	SD	Mean	SD
Number of data samples	13-18	10-15	2	2	4-6	4-6	6-11	6-11	5-7	2-5	5-6	3

Abbreviations: Abbreviations: AP, appetite; CF, cognitive functioning; CO, constipation; DI, diarrhoea; DY, dyspnoea; EF, emotional functioning; FA, fatigue; FI, financial difficulties; GH, global health; MAE, mean absolute error; NV, nausea/vomiting; PA, pain; PF, physical functioning; RF, role functioning; SF, social functioning; SD, standard deviation; SL, sleep/insomnia

*When considering the financial difficulties domain only, the number of patients with standard deviation data was 44-661

Although Kurosawa et al 2016[119] also included more than two time-based health states post-HSCT (<1 year, 1-2 years, 3-5 years, >5 years), the adjusted means presented in the study were judged to have less face validity that those mapped from the Grulke et al 2012[131] study, with patients alive in the first year post-HSCT without GvHD having a significantly lower quality of life (0.51) than patients alive in the first year with GvHD (0.71), and an upper bound reported for the 95% confidence interval outside of the possible range of EQ-5D values for patients alive in the first year with GvHD (1.03).

HSCT-related utility estimates from TA399[102] were excluded from the base case analysis, due to the post-HSCT recovery utility value generated in the original submission (0.771) exceeding the general population norms generated for patients aged 75 (0.758). Similarly, when comparing the utility estimates from Uyl de Groot et al 1998,[100] considered in TA523[91] for HSCT patients, against the UK EQ-5D population norms for the median age of patients in the study (mean estimate not available), the utility estimate of 0.806 exceeded that of a 68 year old patient from the general population utility function (0.793).

Despite the average age of the patient population in Grulke et al 2012[131] (between 40 and 45 years) being considerably younger than the phase III trial population, and unadjusted long-term post-HSCT remission patient utilities being higher than the general population norms for the mean age of patients from the phase III trial (59.6 years), more realistic estimates for health state utility progression were generated following age and sex based adjustment of the utilities. For example, applying the mapping algorithm used for the base case model (Proskorovsky et al 2014[127]) generated utilities for patients aged 65 in long-term remission post-HSCT (Year 4+) of 0.785, slightly below the expected general population utility of 0.807, which was considered the most plausible proportional disutility for functionally cured patients compared to the general population based on review of prior submissions and clinical expert opinion feedback, and to be in alignment with the assumption of excess mortality (e.g. due to late complications) for long-term HSCT remission patients described in Section B.3.3.1.4.

A summary of the unadjusted and age/sex adjusted utility estimates for the baseline age applied in the model (59.6 years) for each mapping algorithm and health state from Grulke et al 2012[131] are described below in Table 49.

Table 49: Summary of mapped EQ-5D utility values from Grulke et al 2012[131]

Health state	Proskorovsky et al 2014 (Full model)		Crott 2010 (Model 3)		McKenzie 2009		Kontodimopoulos 2009		Marriott 2016	
	Unadjusted	Age/sex adjusted*	Unadjusted	Age/sex adjusted*	Unadjusted	Age/sex adjusted*	Unadjusted	Age/sex adjusted*	Unadjusted	Age/sex adjusted*
Induction/HSCT	0.558	0.519	0.613	0.571	0.361	0.336	0.507	0.472	0.583	0.543
Discharge	0.660	0.615	0.691	0.644	0.474	0.442	0.637	0.593	0.647	0.603
Post-HSCT recovery (≤6 months)	0.756	0.704	0.810	0.754	0.611	0.569	0.747	0.695	0.725	0.675
Post-HSCT recovery (7-12 months)	0.818	0.762	0.832	0.775	0.713	0.664	0.827	0.771	0.790	0.736
Post-HSCT recovery (1-3 years)	0.822	0.766	0.826	0.770	0.707	0.658	0.855	0.796	0.785	0.731
Post-HSCT recovery (Year 4+)	0.870	0.810	0.858	0.799	0.746	0.695	0.891	0.830	0.813	0.757
General population utility for patients aged 59 (Ara et al 2010)[159]	0.833									

Abbreviations: HSCT, haematopoietic stem cell transplant

*Adjusted using general population utility function from Ara et al 2010[159]

As shown in Table 49, the Proskorovsky et al 2014[127] mapping algorithm produced one of the two most plausible set of estimates, with three of the other algorithms (Crott et al 2010,[138] McKenzie et al 2009,[126] Marriott et al 2017[140]) producing slightly lower utility for patients 1-3 years post-HSCT than those 7-12 months post-HSCT. Furthermore, as described in Section B.3.3.2.3, the Proskorovsky et al 2014[127] mapping study population was considered most similar to HSCT patients with AML and MDS, and as such these estimates were used in the base case analysis. More conservative estimates of utility improvement from the McKenzie et al 2009,[126] mapping algorithm explored in scenario analysis.

Although non EQ-5D preference-based utility estimates were considered less aligned with the NICE reference case, estimates from Castejon et al 2018[114] were explored in a TTO/DCE scenario analysis, in conjunction with the DCE-based adverse event disutilities from Stein et al 2018.[118] Castejon et al 2018[114]was selected due to provision of more suitable utility estimates than the other TTO, DCE and VAS derived utilities, with data generated from a UK population and values that did not exceed the general population EQ-5D norms from Ara 2010.[159]

In order to achieve internal consistency in the model with the assumption of increased mortality risk for patients in long-term remission post-HSCT compared to the general population, long-term post-HSCT remission utilities (Year 4+) were used for functionally cured patients. For the TTO/DCE utility scenario analysis, the functionally cured health state utility of 0.760 from Castejon et al 2018[114] was applied.

No studies identified in the literature review contained explicit estimates of post-HSCT relapse utilities for AML and MDS patients. In addition to several non EQ-5D preference-based utility studies (Castejon et al 2018[114], Joshi et al 2019,[115] Stein et al 2018,[118] Stein et al 2019,[98] Szende et al 2009,[120] Goss et al 2006[95]), three sources of utility estimates were identified for inclusion in the economic analysis more aligned with the NICE reference case, and with generally more conservative estimates of differentiation in utility between remission and relapse/progression patients. Leunis et al 2014[116] estimated a utility value of 0.780 for relapse patients compared to 0.830 for patients with no relapse, based on Dutch EQ-5D tariffs. Mamolo et al 2017[117] included a mean estimate for relapsed/refractory patients of

0.730, compared to 0.740 and 0.760 for first line AML treatment and untreated AML patients respectively, based on a cross sectional survey of US patients with de novo AML. However, in both cases, the estimates lacked face validity in terms of application to post-HSCT relapse patients, particularly when adjusted for age and sex, with health state utilities generated similar to (or in some cases exceeding) short-term post-HSCT remission estimates.

Two separate mapped EQ-5D estimates of 0.623 and 0.568 based on QLQ-C30 clinical trial data were available from TA399,[102] generated using the Proskorovsky et al 2014[127] and McKenzie et al 2009[126] mapping studies respectively. While these values initially appeared to have improved face validity compared to alternative EQ-5D related estimates, the estimates were from an elderly patient population with a mean age of 75 years, which was significantly older than the treosulfan phase III trial population.

Despite age and sex based utility adjustment, EQ-5D estimates generated for relapse patients were still higher than the estimates from all sources considered for discharge utility estimates, with the exception of the DCE based estimates from Stein et al 2018[118] and Stein et al 2019.[98] In addition, the estimates were also still higher than some sources of post-HSCT recovery utilities considered for short-term post-HSCT recovery (up to one year).

Given the expectation that patients experiencing a relapse within the first 28 days post-HSCT would not be discharged from hospital, and that patients discharged from hospital 28 days post-HSCT would most likely be in a remission/recovery state, it was considered implausible that utility values for relapse patients would be higher than those discharged from hospital or in short-term recovery post-HSCT. Furthermore, clinical expert opinion feedback suggested that utilities for post-HSCT relapse patients would be considerably lower than patients relapsing earlier in the treatment pathway (i.e. pre-HSCT). Under the assumption that the lack of face validity in the estimates was a result of population differences between studies (and reflective of a less severe patient population), and that the absolute or proportional differences in utility from the available sources of relapse data compared to remission or post-HSCT health state utilities presented in those studies would provide a more accurate reflection of utility decrements for patients experiencing a

relapse following HSCT, the ability to apply utility multipliers and disutilities for AML and MDS relapse patients was incorporated into the model. Following clinical expert input, it was considered plausible to apply these utility multipliers and disutilities to patients in short-term HSCT recovery (≤6 months) in order to generate more appropriate health state utility estimates for the base case model.

The Proskorovsky et al 2014[127] based relapse utility multiplier estimated from TA399 was used in the base case economic analysis for internal consistency with the other health state utilities, with the McKenzie et al 2009[126] utility multiplier explored in scenario analysis. As the second clinical expert interviewed indicated that the quality of life of post-HSCT relapse/progression patients may in fact be lower than the estimate considered for the base case model, application of the disutility multiplier to the discharge patient utility was explored as a scenario analysis.

For the TTO/DCE based scenario analysis, the relapse utility estimate from Castejon et al 2018[114] was applied using the standard utility function multiplier calculation used to derive other health state utilities (rather than applying a utility multiplier or disutility vs short-term HSCT recovery patients).

All health state utilities were adjusted according to age and sex using a utility function for general population norms from Ara et al 2010.[160] The general population utility function for the base case analysis was calculated using a female population proportion of 39.2%, based on the overall patient population of the Phase III trial. The coefficients for the general population utility function and estimated general population utility functions are shown below in Table 50 and Figure 19.

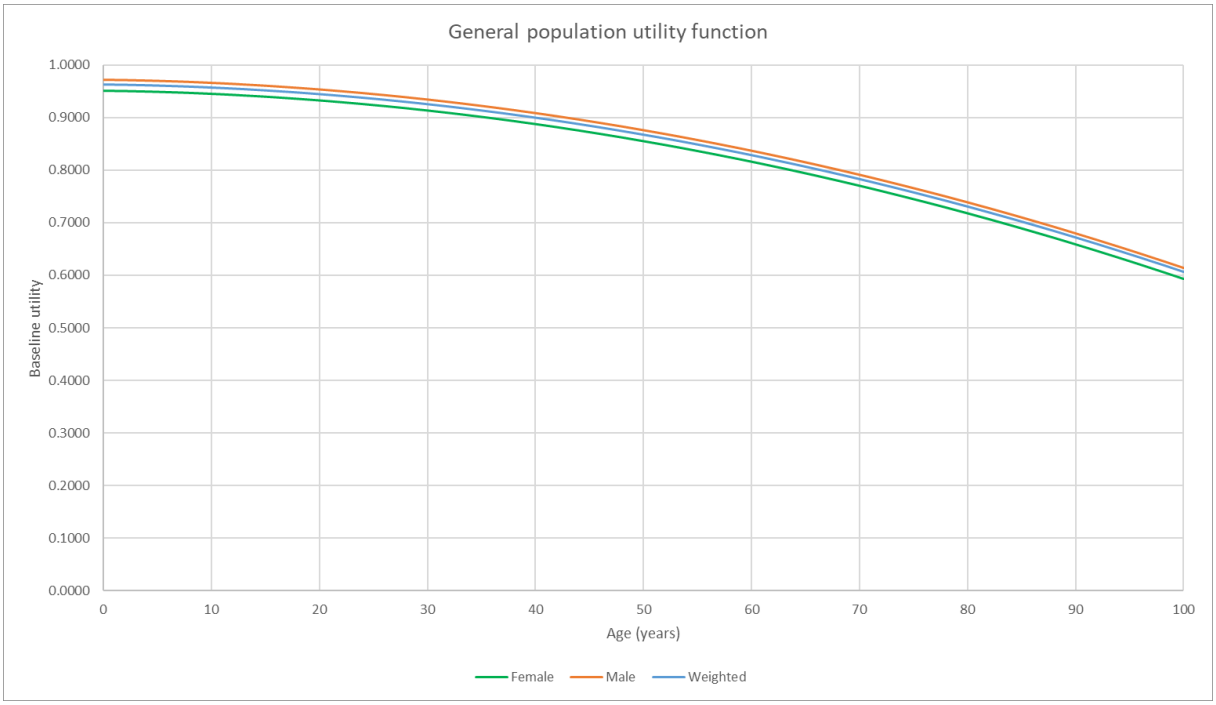
Table 50: Summary of general population EQ-5D utility function coefficients from Ara et al 2010[160]

Variable	Coefficient	SE*
Intercept	0.0212126	0.0042425
Sex	-0.0002587	0.0000517
Age	-0.0000332	0.0000066
Age squared	0.9508566	0.1901713

Abbreviations: OLS, ordinary least squares; UK, United Kingdom

*Assumed to be 20% of the coefficient value due to lack of precision data described in the study

Figure 19: Weighted general population and sex specific utility functions derived from Ara et al 2010[160]



Utility function multipliers and disutilities were then calculated for each health state utility value by comparing it against the general population utility function for the average age of patients within the reference population of the original quality of life study. These multipliers and disutilities were subsequently applied to the general population utility function curve to produce health state utility function curves for ages 0-100 per health state. For the base case model, the multiplier approach was used for both the health state utility functions and for generating the relapse/progression curve in relation to short-term post-HSCT recovery (≤6 months), with slight convergence of health state utilities with increasing age considered to be a more plausible and conservative approach than assuming preservation of absolute differences in utility. Utility estimates for the baseline patient age (59.6 years) are shown in Table 51, with the utility functions for the base case model and scenario analyses are shown in Figure 20,

Figure 21, Figure 22, and

Figure 23. Each figure shows an example progression of a long-term HSCT recovery patient remaining in remission until age 100 (black dashed line) using the baseline age of 59.6 years.

Table 51: Summary of utility values used in the base case model and scenario analyses

Health state	Base case utilities		Utility scenario analysis 1		Utility scenario analysis 2		Utility scenario analysis 3	
	Utility values*	Source	Utility values*	Source	Utility values*	Source	Utility values*	Source
Induction/HSCT	0.519	Grulke et al 2012[131] (Proskorovsky et al 2014 mapping)	0.336	Grulke et al 2012[131] (McKenzie et al 2009 mapping)	0.280	Castejon et al 2018[114]	0.519	Grulke et al 2012[131] (Proskorovsky et al 2014 mapping)
Discharge	0.615		0.442		0.620		0.615	
Post-HSCT recovery (≤6 months)	0.704		0.569		0.620		0.704	
Post-HSCT recovery (7-12 months)	0.762		0.664		0.620		0.762	
Post-HSCT recovery (Year 2 and 3)	0.766		0.658		0.620		0.766	
Post-HSCT recovery (Year 4+)	0.810		0.695		0.620		0.810	
Functionally cured	0.810		0.695		0.760		0.810	
Relapse/progression (AML)	0.569†	TA399[102] (Proskorovsky et al 2014 mapping)	0.436†	TA399[102] (McKenzie et al 2009 mapping)	0.100		0.569†	TA399[102] (Proskorovsky et al 2014 mapping)
Relapse/progression (MDS)	0.569†		0.436†		0.100		0.538†	Szende et al 2009[120]
Extensive cGvHD disutility	-0.120	Kurosawa et al 2016[119]	-0.120	Kurosawa et al 2016[119]	-0.190		-0.120	Kurosawa et al 2016[119]
Stage III/IV aGvHD disutility	-0.120		-0.120		-0.190		-0.120	
Sepsis/lung infection	-0.024	TA399[102] (Proskorovsky et al 2014 mapping)	-0.021	TA399[102] (Mckenzie et al 2009 mapping)	-0.218	Stein et al 2018[118]	-0.024	TA399[102] (Proskorovsky et al 2014 mapping)
Diarrhoea	-0.024		-0.021		-0.176		-0.024	
Oral mucositis and maculopapular rash	-0.024		-0.021		-0.060		-0.024	

Health state	Base case utilities		Utility scenario analysis 1		Utility scenario analysis 2		Utility scenario analysis 3	
	Utility values*	Source	Utility values*	Source	Utility values*	Source	Utility values*	Source
Febrile neutropenia	-0.024		-0.021		-0.100		-0.024	
All other Grade 3+ CTCAEs	-0.024		-0.021		-0.024	TA399[102] (Proskorovsky 2014 mapping)	-0.024	

Abbreviations: CTCAE, common terminology criteria for adverse events; HSCT, haematopoietic stem cell transplant

*Health state utilities are adjusted using a general population utility function derived from Ara et al 2010†[160]

†Generated using a disutility multiplier applied to short-term HSCT recovery patients (≤6 months)

Figure 20: Age and sex adjusted base case utility functions

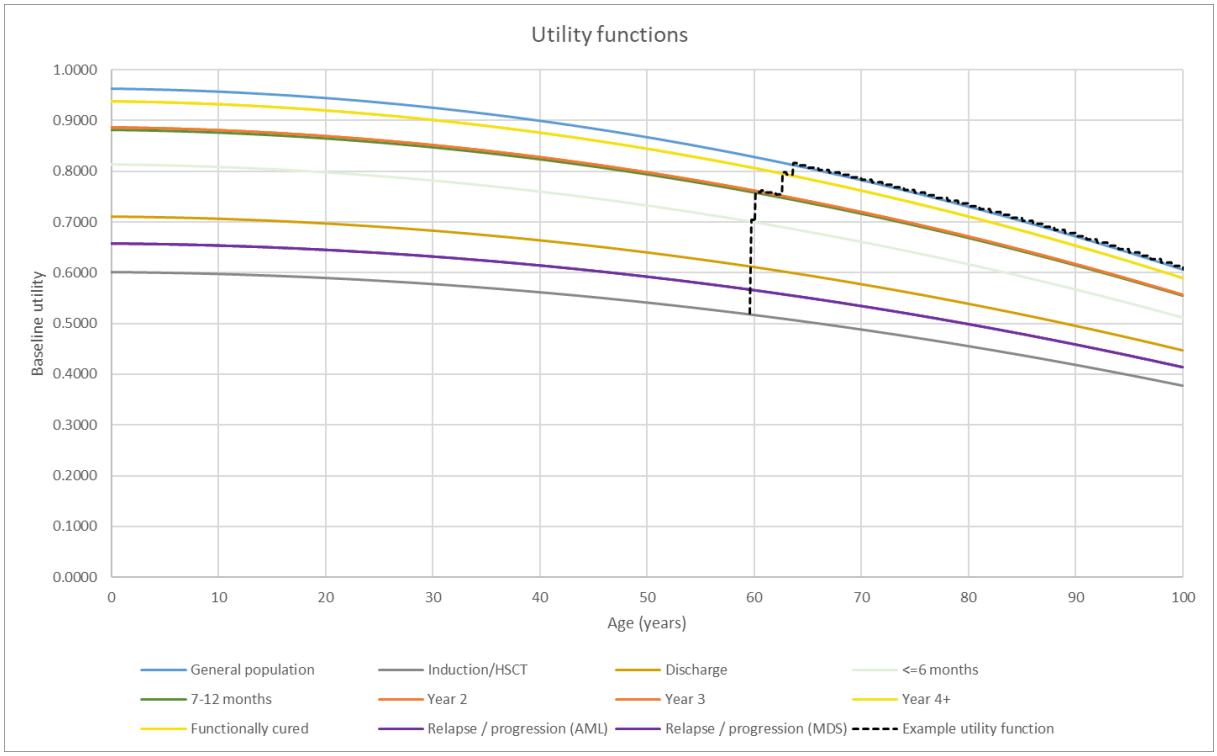


Figure 21: Age and sex adjusted scenario analysis 1 utility functions

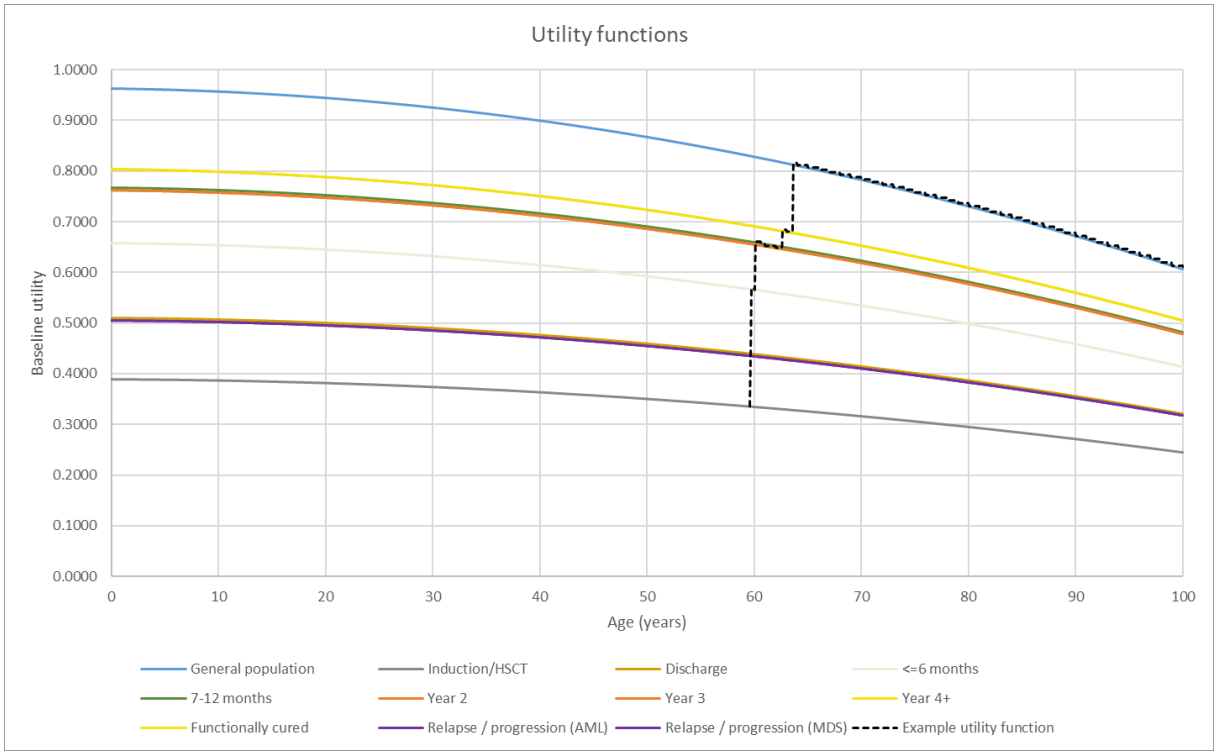


Figure 22: Age and sex adjusted scenario analysis 2 utility functions

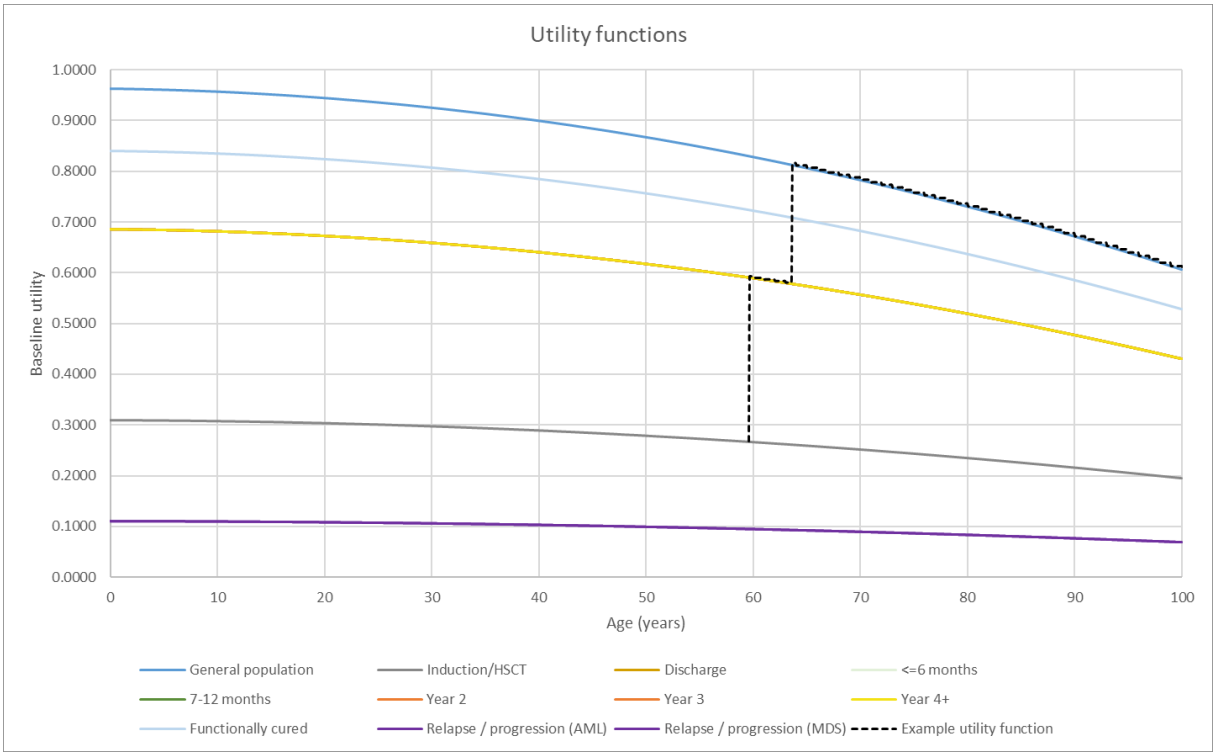
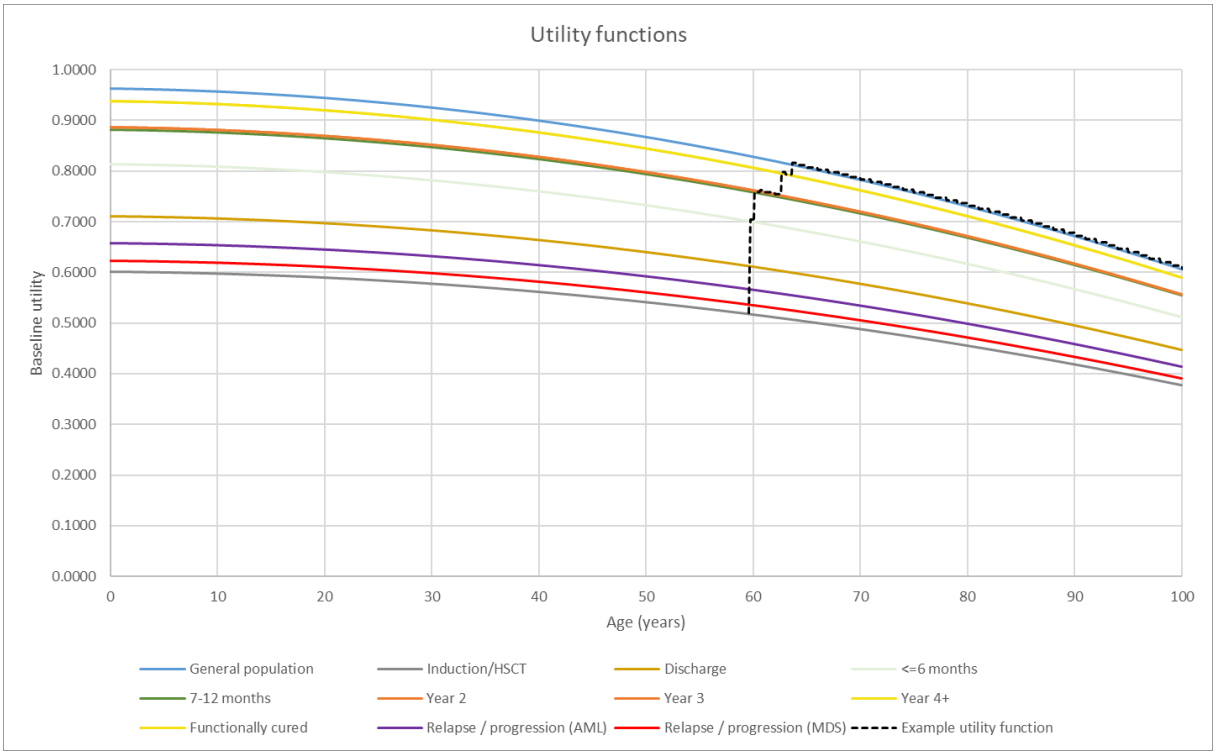


Figure 23: Age and sex adjusted scenario analysis 3 utility functions



One study considered for the cost-effectiveness model (Joshi 2019[115]) provided a negative utility estimate of -0.210 for patients receive a stem cell transplant procedure. Although this study was not used in the base case analysis or key

scenario analyses, the utility function estimation was adjusted such that any negative utility values applied in the model in conjunction with the age and sex adjusted utility function multiplier approach were assumed to be constant with increasing or decreasing age, in order to prevent implausible extrapolation of negative utility values.

B.3.4 Cost and healthcare resource use identification, measurement and valuation

B.3.4.1 Intervention and comparators' costs and resource use

Treatment costs were applied for intravenously administered treosulfan and busulfan based on the regimens received by patients in the trial. Concomitant treatments used in busulfan patients were phenytoin, fludarabine, anti-thymocyte globulin (ATG; MUD patients only), ciclosporin (i.v. and oral), methotrexate and calcium folinate. Excluding phenytoin, the same therapies were used concomitantly with treosulfan patients. Administration costs were excluded from the model, as there was no difference in the administration time for treosulfan and busulfan in the clinical trial (120 minutes), and it was assumed that they were captured as part of the inpatient costs estimated for the HSCT procedure.

All therapies except phenytoin were dosed according to weight or body surface area (BSA). For the base case model, a mean weight of 80.2kg and a mean BSA of 1.931m^2 were used based on the overall phase III trial population.

busulfan treatment costs were based on the dosing administered in the clinical trial (4 x 0.8 mg/kg/day), with a single 3.2mg/kg daily dose applied in a scenario analysis.

Treatment costs were sourced from the British National Formulary (BNF).[161] For all concomitant treatments, dosing was based on the least costly pack costs per mg, with no wastage costs applied (as the only additional treatment for busulfan patients was orally administered).

HSCT procedure costs were estimated from NHS reference costs,[162] using Healthcare Resource Group (HRG) codes for bone marrow and peripheral cell harvest costs, and HSCT procedure codes for adult patients (age 19 and over). A

weighted average of the harvesting and HSCT procedure costs was used to calculate an average HSCT cost for each treatment arm, with the average overall cost for treosulfan and busulfan applied for secondary HSCTs.

For treosulfan and busulfan, wastage costs were applied, with 100% vial wastage assumed in the base case model based on clinical expert opinion.

Treosulfan and busulfan had similar overall intervention costs, with a slightly lower overall intervention cost for treosulfan in the base case analysis due to vial wastage. Treatment costs and HSCT procedure costs are summarised in Table 52.

Table 52: Summary of treatment and HSCT procedure costs

Concomitant therapy	Method of administration	Dosing regimen	Unit cost (per vial/tablet)	Cost per mg	Total treatment cost (including wastage)	Notes/Comments
Treosulfan	i.v.	10 g/m^2 at days -4 to -2 prior to alloHSCT (within 120 minute period)	£53.83 (1g vial) and £208.03 (5g vial)	£0.05 (1g vial) and £0.04 (5 vial)	£2,496.41	Total cost excluding wastage: £2,410.28
Busulfan	i.v.	4 x 0.8 mg/kg/day at days -4 and -3 prior to alloHSCT (within 120 minute period)	£191.19	£3.19	£3,059.00	Total cost excluding wastage: £1,635.55
Fludarabine	i.v.	30 mg/m^2 at days -6 to -2 prior to alloHSCT (within 30 minute period)	£147.07	£2.94	£851.96	Based on price for Fludara 50mg powder for solution for injection vials
ATG	i.v.	10 mg/kg at days -4 to -2 prior to alloHSCT (Grafalon) or 2.5 mg/kg at days -2 and -1 prior to alloHSCT (Thymoglobuline)	£158.77	£6.35	£1,945.82	No BNF price available for Grafalon – treatment costs/dosing based on Thymoglubuline Weighted based on proportion of alloHSCTs with MUDs (76.4%)
Ciclosporin	i.v.	3 mg/kg/day i.v. at start (day -1 before and day of alloHSCT)	£11.01	£0.04	£21.18	Based on price for Sandimmun 250mg/5ml concentrate for solution for infusion ampoules
Ciclosporin	Oral	5 mg/kg/day oral (days +1, +3 and +6 after alloHSCT)	£2.28	£0.02	£27.38	Based on 100mg capsule Drug Tariff (Part VIIIA Category C) price
Methotrexate	i.v.	15 mg/m^2/day (dose 1 at day +1 after alloHSCT) then 10 mg/m^2/day (doses 2 and 3 at days +3 and +6 after alloHSCT)	£200.57	£0.04	£2.71	Based on price for 5g/200ml solution for infusion vials
CA-Folinate	i.v.	15 mg/m^2 (dose 1 at day +1 after alloHSCT) then 10 mg/m^2/day (doses 2 and 3 at days +3 and +6 after alloHSCT)	£4.62	£0.15	£10.42	Based on price for Refolinon 30mg/10ml solution for injection ampoules

Concomitant therapy	Method of administration	Dosing regimen	Unit cost (per vial/tablet)	Cost per mg	Total treatment cost (including wastage)	Notes/Comments
Phenytoin	Oral	3 x 200 mg at day -5 before alloHSCT, then 3 x 100 mg at days -4 to -2 prior to alloHSCT	£0.33	£0.001	£3.14	Based on 300mg capsule Drug Tariff (Part VIIIA Category A) price Applied only to busulfan patients
HSCT procedure	-	-	-	-	£40,774.35	Calculated as the sum of the weighted average of harvesting costs (HRG codes SA18Z and SA34Z) and allogeneic transplant costs (HRG codes SA20A-SA23A, SA38A, SA39A, SA40Z) for elective inpatients
Total (treosulfan)	**£46,130.24**					
Total (busulfan)	**£46,695.97**					

Abbreviations: alloHSCT, allogeneic haematopoietic stem cell transplant; ATG, anti-thymocyte globulin; i.v., intravenous; MUD, matched unrelated donor; N/A, not applicable

B.3.4.2 Adverse reaction unit costs and resource use

In addition to extensive cGvHD and stage III/IV aGvHD, all treatment related Grade 3+ Common Terminology Criteria for Adverse Events (CTCAEs) with an incidence ≥1% were included in the model. As cGvHD cannot occur until day 100, incidence estimates below reflect events captured between day 100 and day 731 in the trial, with aGvHD events captured for the first 100 days, and all other AEs for the first 28 days. Associated disutilities and costs were then estimated by weighting the overall disutility and cost by the adverse event incidence.

Adverse event incidence is summarised below in Table 53.

Table 53: Summary of treatment related adverse events with ≥1% incidence from the phase III clinical trial

Adverse event	Cumulative incidence	
	Busulfan	Treosulfan
Extensive cGvHD (days 100-731)	26.7%	19.7%
Stage III/IV aGvHD (up to 100 days)	8.1%	6.4%
Mucositis oral	6.0%	4.4%
Nausea	4.9%	2.6%
Diarrhoea	0.7%	1.1%
Vomiting	1.4%	0.4%
Gamma glutamyl transferase (GGT) increased	8.1%	3.0%
Alanine aminotransferase increased	2.8%	4.8%
Aspartate aminotransferase increased	2.1%	4.1%
Blood bilirubin increased	1.4%	2.2%
Investigations (other)	1.4%	1.5%
Febrile neutropenia	4.9%	4.4%
Sepsis	0.4%	2.2%
Lung infection	0.7%	1.1%
Anorexia	1.4%	1.5%
Syncope	1.4%	0.0%
Rash maculo-papular	1.1%	0.7%

Abbreviations: aGvHD, acute graft vs host disease; cGvHD, chronic graft vs host disease; GGT, gamma glutamyl transferase

In order to apply more accurate discounting of utilities and costs for adverse events, mean duration estimates were used to spread AE disutilities and costs over multiple cycles, assuming a constant rate of incidence over time. Incidence start cycle and incidence end cycle weights were estimated to partition incidence estimates for GvHD events that were only possible to occur for a particular time frame within a 28 day cycle.

For extensive cGvHD and Stage III/IV acute GvHD, mean duration estimates of 9 months and 2.5 months were used respectively, based on TA545,[103] and validated through clinical expert opinion. Other adverse events sourced via NHS reference cost data were assumed to have an event duration equal to the average inpatient stay for the associated HRG codes. For events costed using NHS Tariff data (due to lack of appropriate codes in the latest NHS reference costs) and maculopapular rash, event durations were based on mean estimates from the phase III clinical trial, with analysis performed in Stata 14.

Limited data were identified for GvHD costs in the literature. Two French studies (Robin et al 2017[163] and Esperou et al 2004[164]) and one US study (Khera et al 2014[165]) were identified through targeted review, with Robin et al 2017[163] and Khera et al 2014[165] included in the model. Despite its inclusion in a prior AML technology appraisal (TA545[103]), the Esperou et al 2004[164] study was excluded due to concerns around the accuracy of uplifting costs over a 17 year time period (from 2001 Euros). Robin et al 2017[163] was used for the base case analysis as French costs were considered more generalisable to a UK population. Costs were converted from 2015 Euros to GBP using a purchasing power parity estimate (PPP) from the Organisation for Economic Co-operation and Development (OECD)[166], and inflated to 2018 costs using the Personal Social Services Research Unit (PSSRU)[167] Hospital and Community Health Services (HCHS) pay and prices index. In order to prevent potential overestimation of GvHD costs from Robin et al 2017[163], excess costs for cGvHD and stage II-IV aGvHD (i.e. the difference between groups with/without cGvHD, and patients with stage 0/I aGvHD and stage II-IV aGvHD) were applied.

For all other adverse events included in the model, NHS reference costs were used as a primary source of duration and cost data, with prior technology appraisals used to help select appropriate HRG codes. Elective inpatient codes were used based on

the assumption that as these adverse events were observed within the first 28 days post-HSCT and the average duration of an HSCT procedure inpatient stay (based on the HRG codes for elective inpatients) was around 27 days that these adverse events would likely be treated within the same episode of inpatient care.

However, for costing investigations (such as increased GGT) and febrile neutropenia, HRG codes used in TA523[91] were not available since the 2014/2015 version of the NHS reference costs.[162] Therefore, the most recent version of the NHS Tariffs[168] including these HRG codes (2016/2017) were considered a more appropriate source of event costs. Elective inpatient costs were unavailable from 2016/2017 tariffs – as such, non-elective inpatient costs were used.

For maculopapular rash events, clinical expert opinion confirmed that patients would not require an extended length of stay in hospital, and that patients would most likely receive treatment with systemic steroids. As such, maculopapular rash events were costed assuming a 10mg daily dose of oral prednisolone and an event duration of 16 days based on the phase III clinical trial.

Adverse event duration and event costs are summarised in Table 54.

Table 54: Summary of cost of treatment related adverse events with ≥1% incidence from the phase III clinical trial

Adverse event	Duration, days		Event cost	
	Mean (SE)	Source	Mean (SE)	Source
Extensive cGvHD	273.9 (54.8)*	Assumption based on TA545[103] and clinical expert opinion	£5,823.68 (N/A)†	Robin et al 2017[163]
Stage III/IV aGvHD	76.1 (15.2)*	Assumption based on TA545 and clinical expert opinion	£14,793.04 (N/A)†	Robin et al 2017
Mucositis oral	2.2 (0.4)*	NHS reference costs 2017/18[104] Weighted average of elective inpatient currency codes CB02A-CB02F (Non-Malignant, Ear, Nose, Mouth, Throat or Neck Disorders, with Interventions).	£2,021.36 (£404.27)*	NHS reference costs 2017/18 Weighted average of elective inpatient currency codes CB02A-CB02F (Non-Malignant, Ear, Nose, Mouth, Throat or Neck Disorders, with Interventions).
Nausea	3.3 (0.7)*	NHS reference costs 2017/18 Weighted average of	£2,263.78 (£452.76)*	NHS reference costs 2017/18 Weighted average of elective

Adverse event	Duration, days		Event cost	
	Mean (SE)	Source	Mean (SE)	Source
		elective inpatient currency codes FD10A-FD10M (Non-Malignant Gastrointestinal Tract Disorders).		inpatient currency codes FD10A-FD10M (Non-Malignant Gastrointestinal Tract Disorders).
Diarrhoea	3.3 (0.7)*	NHS reference costs 2017/18 Weighted average of elective inpatient currency codes FD10A-FD10M (Non-Malignant Gastrointestinal Tract Disorders).	£2,263.78 (£452.76)*	NHS reference costs 2017/18 Weighted average of elective inpatient currency codes FD10A-FD10M (Non-Malignant Gastrointestinal Tract Disorders).
Vomiting	3.3 (0.7)*	NHS reference costs 2017/18 Weighted average of elective inpatient currency codes FD10A-FD10M (Non-Malignant Gastrointestinal Tract Disorders).	£2,263.78 (£452.76)*	NHS reference costs 2017/18 Weighted average of elective inpatient currency codes FD10A-FD10M (Non-Malignant Gastrointestinal Tract Disorders).
GGT increased	40 (4.7)	Analysis of phase III trial data	£983.58 (£196.72)*	NHS National Tariff Payment System 2016/17[168] PA48B Blood Cell Disorders without CC (Admitted Patient Care & Outpatient Procedures): Non-elective spell tariff Inflated to 2017/18 using PSSRU[167] HCHS index
Alanine aminotransferase increased	15.1 (4.2)	Analysis of phase III trial data	£983.58 (£196.72)*	NHS National Tariff Payment System 2016/17 PA48B Blood Cell Disorders without CC (Admitted Patient Care & Outpatient Procedures): Non-elective spell tariff. In line with Tremblay et al 2018 and TA523 (HRG code not in 2017/18 tariff) Inflated to 2017/18 using PSSRU HCHS index
Aspartate aminotransferase increased	8.7 (1.8)	Analysis of phase III trial data	£983.58 (£196.72)*	NHS National Tariff Payment System 2016/17 PA48B Blood Cell Disorders without CC (Admitted Patient Care & Outpatient Procedures): Non-elective spell tariff.

Adverse event	Duration, days		Event cost	
	Mean (SE)	Source	Mean (SE)	Source
				Inflated to 2017/18 using PSSRU HCHS index
Blood bilirubin increased	23.4 (8.1)	Analysis of phase III trial data	£983.58 (£196.72)*	NHS National Tariff Payment System 2016/17 PA48B Blood Cell Disorders without CC (Admitted Patient Care & Outpatient Procedures): Non-elective spell tariff. Inflated to 2017/18 using PSSRU HCHS index
Investigations (other)	31.3 (11.8)	Analysis of phase III trial data	£983.58 (£196.72)*	NHS National Tariff Payment System 2016/17 PA48B Blood Cell Disorders without CC (Admitted Patient Care & Outpatient Procedures): Non-elective spell tariff. Inflated to 2017/18 using PSSRU HCHS index
Febrile neutropenia	4.1 (1.1)	Analysis of phase III trial data	£3,669.67 (£733.93)*	NHS National Tariff Payment System 2016/17. PA45Z Febrile Neutropenia with Malignancy (Admitted Patient Care & Outpatient Procedures): Non-elective spell tariff. Inflated to 2017/18 using PSSRU HCHS index
Sepsis	6.1 (1.2)*	NHS reference costs 2017/18 Weighted average of elective inpatient currency codes WJ06A-WJ06J (Sepsis).	£3,662.50 (£732.50)*	NHS reference costs 2017/18 Weighted average of elective inpatient currency codes WJ06A-WJ06J (Sepsis).
Lung infection	5.4 (1.1)*	NHS reference costs 2017/18 Weighted average of elective inpatient currency codes DZ11K-N;P-V (Lobar, Atypical or Viral Pneumonia).	£2,924.23 (£584.85)*	NHS reference costs 2017/18 Weighted average of elective inpatient currency codes DZ11K-N;P-V (Lobar, Atypical or Viral Pneumonia).
Anorexia	3.2 (0.6)*	NHS reference costs 2017/18 Weighted average of elective inpatient currency	£2,036.45 (£407.29)*	NHS reference costs 2017/18 Weighted average of elective inpatient currency codes

Adverse event	Duration, days		Event cost	
	Mean (SE)	Source	Mean (SE)	Source
		codes FD04A-E (Nutritional Disorders).		FD04A-E (Nutritional Disorders).
Syncope	2 (0.4)*	NHS reference costs 2017/18 Weighted average of elective inpatient currency codes EB08A-E (Syncope or Collapse).	£1,256.05 (£251.21)*	NHS reference costs 2017/18 Weighted average of elective inpatient currency codes EB08A-E (Syncope or Collapse).
Rash maculo-papular	16 (6)	Analysis of phase III trial data	£21.44 (£4.29)*	BNF[161] Assumption of 10mg daily dose of oral prednisolone.

Abbreviations: aGvHD, acute graft vs host disease; BNF, British National Formulary; cGvHD, chronic graft vs host disease; GGT, gamma glutamyl transferase; HCHS, Hospital and Community Health Services; N/A, not applicable; NHS, National Health Service; SE, standard error.
*SE assumed to be 20% of the mean
†Not applicable as these estimates were derived as the difference in costs for patients with/without cGvHD and with stage 0/I and stage II-IV aGvHD respectively
Source: TA545[103]; Robin et al 2017[163]; NHS reference costs 2017/18;[162] NHS National Tariff Payment System 2016/17[168]; PSSRU[167] and BNF[161]

B.3.4.3 Miscellaneous unit costs and resource use

No additional miscellaneous costs were included.

B.3.5 Base-case results

B.3.5.1 Base-case incremental cost-effectiveness analysis results

B.3.5.1.1 Base case deterministic results

The base-case deterministic results of the analysis are presented below in Table 55. Treosulfan was dominant over busulfan, with lower total costs and higher total QALYs.

Table 55: Base-case cost-effectiveness results

Technologies	Total costs (£)	Total LYG	Total QALYs	Incremental costs (£)	Incremental LYG	Incremental QALYs	ICER versus baseline (£/QALY)	ICER incremental (£/QALY)
Treosulfan	£137,062	8.73	6.44					
Busulfan	£160,821	7.71	5.55	£23,759	-1.01	-0.89	Dominated	Dominated

Abbreviations: ICER, incremental cost-effectiveness ratio; LYG, life years gained; QALYs, quality-adjusted life-years

B.3.5.1.2 Results by health state

Disaggregated results by health state are presented below in Table 56.

Table 56: Disaggregated results by health state

Health state	Costs		QALYs		LYG (undiscounted)	
	Treosulfan	Busulfan	Treosulfan	Busulfan	Treosulfan	Busulfan
Event-free survival	£102,989.86	£100,798.67	6.137	4.967	11.323	9.132
Relapse / progression: AML	£15,898.11	£31,800.17	0.193	0.372	0.476	0.965
Relapse / progression: MDS	£9,468.31	£18,909.51	0.109	0.210	0.269	0.546
Dead	£8,705.45	£9,312.29	-0.004	-0.007	12.069	10.643
Total	£137,061.73	£160,820.64	6.435	5.549	8.726	7.711

Abbreviations: AML, acute myeloid leukaemia; LYG, life year gain; MDS, myelodysplastic syndrome; QALYs, quality-adjusted life-year.

B.3.6 Sensitivity analyses

B.3.6.1 Probabilistic sensitivity analysis (PSA)

Probabilistic sensitivity analysis was performed with 5,000 simulations to capture stochastic uncertainty around each of the model inputs, with simultaneous variation for each parameter according to a predefined probability distribution. For inputs with no uncertainty data available, standard errors were estimated as 20% of the mean value. For all other inputs, standard errors were taken directly from the associated source, or derived from other measures presented (standard deviations, 95% CIs, etc.). A full list of inputs, associated standard errors and selected distributions are provided in Appendix L (model input list), as well as survival parameter values and Cholesky matrices for all survival functions and patients subgroups has been provided (Appendix M2, M3 and M4).

Results of the probabilistic sensitivity analysis for the base case model are shown in Figure 24 (cost-effectiveness plane) and

Figure 25 (cost-effectiveness acceptability curve; CEAC).

Figure 24: Base case PSA results – cost-effectiveness plane

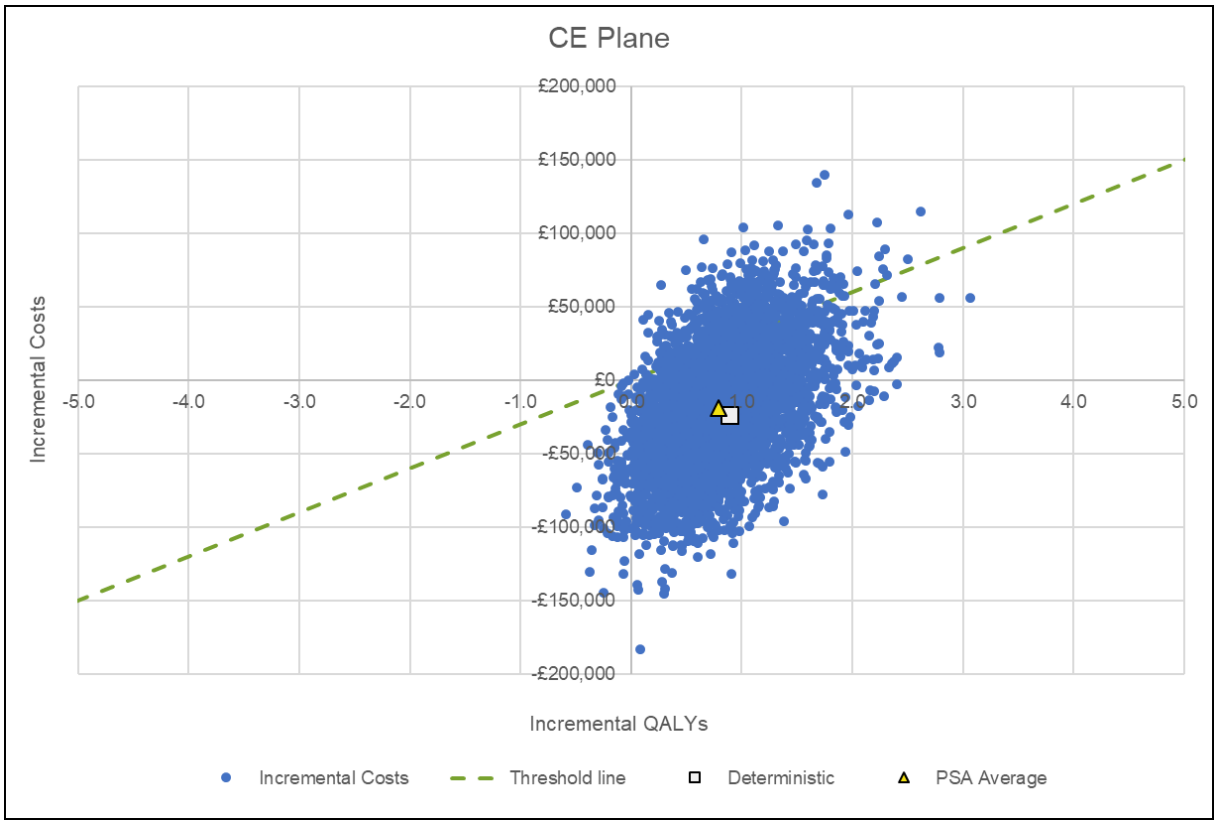
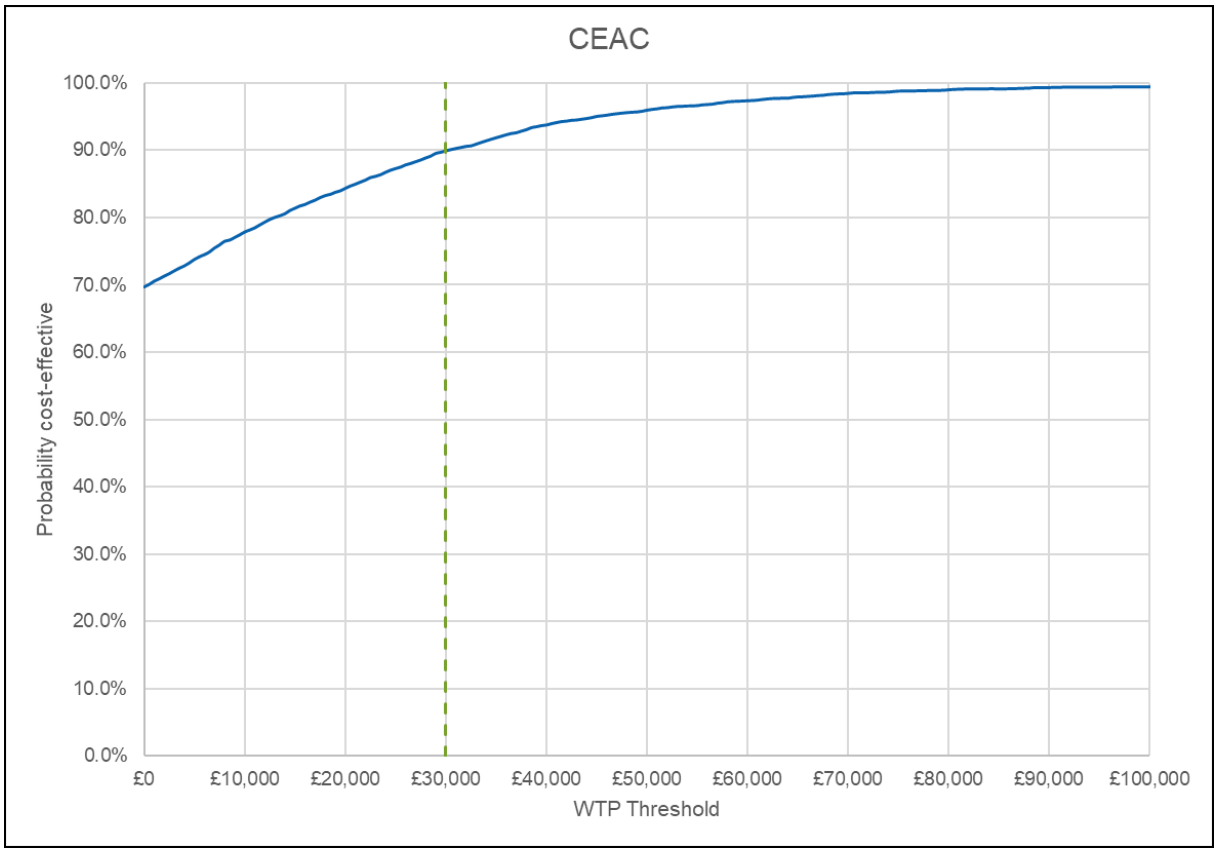


Figure 25: Base case PSA results – cost-effectiveness acceptability curve (CEAC)



The mean probabilistic incremental total costs were -£19,084 for treosulfan compared to busulfan, with mean probabilistic incremental total QALYs of 0.79. The mean incremental total life years were 1.08 for treosulfan compared to busulfan. The mean probabilistic ICER was -£24,118 (95% CI: -£21,481 to -£16,688 for incremental costs, 0.76-0.82 for incremental QALYs), with a mean INMB of £42,824 (based on a £30,000 per QALY threshold).

At a willingness to pay (WTP) threshold of £30,000 per QALY, the probability of treosulfan being cost-effective was 89.9%. Treosulfan was highly cost-effective at all thresholds examined, with a cost-effectiveness probability of 84.3% at a £20,000 per QALY and 69.6% probability at a £0 threshold.

B.3.6.2 Deterministic sensitivity analysis

Univariate sensitivity analysis was also performed, with each parameter varied by +/-20%. Results for the incremental cost-effectiveness ratio (ICER) are shown in Figure 26, with results for the incremental net monetary benefit (INMB) (using a £30,000 per QALY WTP threshold) shown in Figure 27.

The tornado diagrams indicate that the most sensitive inputs in the model were the meanlog coefficients for the NMCM lognormal survival functions for each treatment.

Figure 26: Base case DSA results – ICER

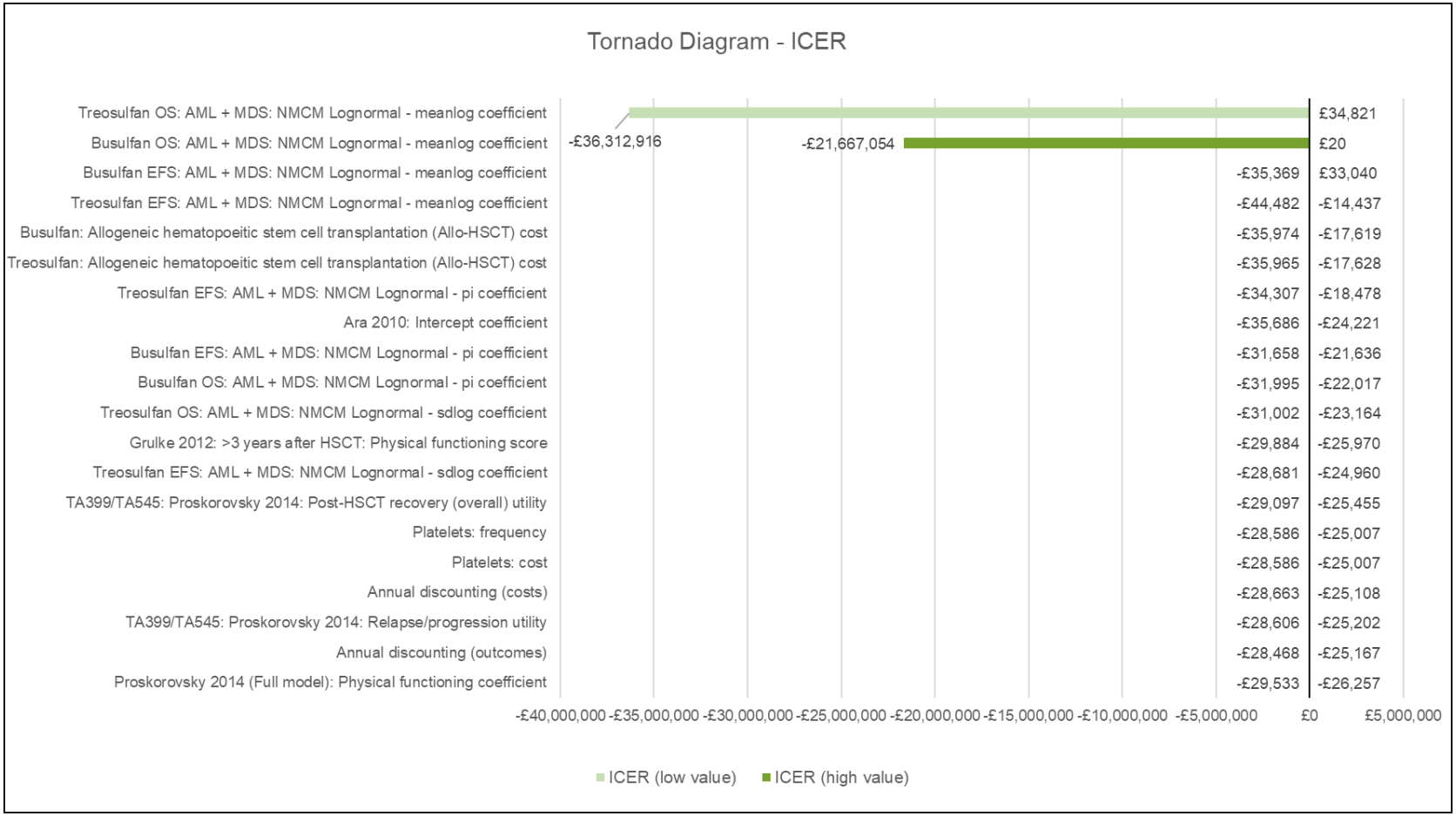
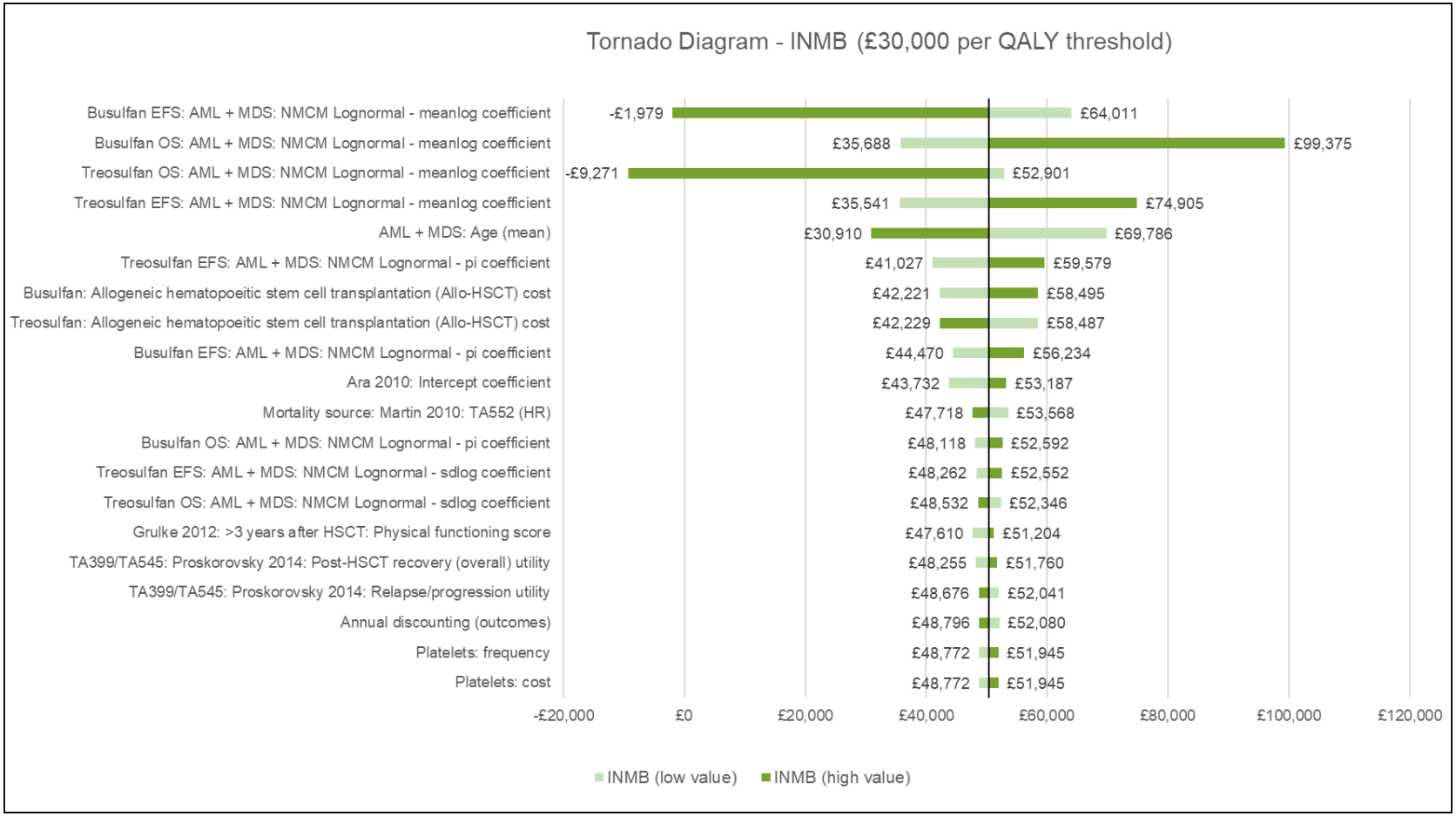


Figure 27: Base case DSA results – INMB (£30,000 per QALY threshold)



B.3.6.3 Scenario analysis

B.3.6.3.1 Scenario 1: Use of McKenzie et al 2009 mapping estimates

Three scenario analyses regarding utility values were performed. The first scenario was to utilise an alternative mapping (McKenzie et al 2009[126]) of the utility values estimated from Grulke et al 2012[131] and those used in TA399[102]. The age and sex adjusted utility parameters (for the baseline age of 59.6 years) are shown below in Table 57.

Table 57: Scenario 1: Use of McKenzie et al 2009 (for Grulke et al 2012 and TA399) utilities

Variable	Base case value (Proskorovsky et al 2014[127] mapping)	Updated scenario value (McKenzie et al 2009 mapping)[126]
Induction / HSCT utility	0.517	0.334
Post-HSCT recovery (short term) discharge	0.612	0.439
Post-HSCT recovery (short term) ≤6 months	0.700	0.566
Post-HSCT recovery (short term) 7-12 months	0.758	0.660
Post-HSCT recovery (long-term) year 2 and year 3	0.762	0.655
Post-HSCT recovery (long-term) year 4+	0.806	0.691
Functionally cured	0.806	0.691
Relapse / progression (AML and MDS patients)*†	0.566	0.434
Grade III+ AEs*	-0.024	-0.021

Source: McKenzie et al 2009,[126] Grulke et al 2012,[131] TA399,[102] and Proskorovsky et al 2014.[127]
Abbreviations: AE, adverse event; AML, acute myeloid leukaemia; HSCT, hematopoietic stem cell transplant; MDS, myelodysplastic syndrome
*Value reported directly in TA399.
†Estimated by applying a utility multiplier vs short-term post-HSCT recovery (≤6 months).

Deterministic results for the scenario analysis are shown below in Table 58.

Table 58: Scenario 1: Cost-effectiveness results

Technologies	Total costs (£)	Total LYG	Total QALYs	Incremental costs (£)	Incremental LYG	Incremental QALYs	ICER versus baseline (£/QALY)	ICER incremental (£/QALY)
Treosulfan	£137,062	8.73	5.47					
Busulfan	£160,821	7.71	4.68	£23,759	-1.01	-0.79	Dominated	Dominated

Abbreviations: ICER, incremental cost-effectiveness ratio; LYG, life years gained; QALYs, quality-adjusted life years

B.3.6.3.2 Scenario 2: Use of Castejon et al 2018 and Stein et al 2018 utility values

For the second scenario analysis, the base case HSUVs were replaced with those from Castejon et al 2018,[114] with serious AE disutilities utilised from Stein et al 2018.[118] As described in Section B.3.3.2.5, the relapse/progression utilities were applied using the standard age-adjusted utility function, rather than as a disutility multiplier applied to short-term post-HSCT recovery patients (≤6 months) as per the base case model. A summary of the age-adjusted utilities for the baseline age and scenario analysis results are shown in Table 59and Table 60 respectively.

Table 59: Scenario 2: Castejon et al 2018[114]/Stein et al 2018[118] utilities

Variable	Base case value	Updated scenario value (Castejon et al 2018[114]/Stein et al 2018[118])
Induction / HSCT utility	0.517	0.266
Post-HSCT recovery (short term) discharge	0.612	0.589
Post-HSCT recovery (short term) ≤6 months	0.700	0.589
Post-HSCT recovery (short term) 7-12 months	0.758	0.589
Post-HSCT recovery (long-term) year 2 and year 3	0.762	0.589
Post-HSCT recovery (long-term) year 4+	0.806	0.589
Functionally cured	0.806	0.722
Relapse / progression (AML and MDS)	0.566	0.095
Extensive chronic GvHD and stage III/IV acute GvHD disutility	-0.120	-0.190
Sepsis/lung infection	-0.024	-0.218

Variable	Base case value	Updated scenario value (Castejon et al 2018[114]/Stein et al 2018[118])
Diarrhoea	-0.024	-0.176
Oral mucositis and maculopapular rash	-0.024	-0.060
Febrile neutropenia	-0.024	-0.100

Abbreviations: AML, acute myeloid leukaemia; GvHD, graft versus host disease; HSCT, hematopoietic stem cell transplant; MDS, myelodysplastic syndrome

Table 60: Scenario 2: Cost-effectiveness results

Technologies	Total costs (£)	Total LYG	Total QALYs	Incremental costs (£)	Incremental LYG	Incremental QALYs	ICER versus baseline (£/QALY)	ICER incremental (£/QALY)
Treosulfan	£137,062	8.73	5.29					
Busulfan	£160,821	7.71	4.33	£23,759	-1.01	-0.96	Dominated	Dominated

Abbreviations: ICER, incremental cost-effectiveness ratio; LYG, life years gained; QALYs, quality-adjusted life years

B.3.6.3.3 Scenario 3: Alternative MDS relapse/progression utility

For the third scenario analysis, an MDS specific utility for RBC transfusion dependent patients (from Szende et al 2009[120]) was applied to MDS relapse patients. Following application of the utility multiplier vs short-term HSCT-recovery (≤6 months), the age-adjusted utility for patients aged 59.6 years was 0.535; results from this analysis are shown below in Table 61.

Table 61: Scenario 3: Cost-effectiveness results

Technologies	Total costs (£)	Total LYG	Total QALYs	Incremental costs (£)	Incremental LYG	Incremental QALYs	ICER versus baseline (£/QALY)	ICER incremental (£/QALY)
Treosulfan	£137,062	8.73	6.43					
Busulfan	£160,821	7.71	5.54	£23,759	-1.01	-0.89	Dominated	Dominated

Abbreviations: ICER, incremental cost-effectiveness ratio; LYG, life years gained; QALYs, quality-adjusted life years

B.3.6.3.4 Scenario 4: Application of relapse utility multiplier to discharge utility

As described in Section B.3.3.2.5, an additional utility scenario was explored where the relapse disutility multiplier from the Proskorovsky et al 2014[127] mapping of TA399 utilities was applied to discharge patients. Results are presented below in Table 62.

Table 62: Scenario 4: Cost-effectiveness results

Technologies	Total costs (£)	Total LYG	Total QALYs	Incremental costs (£)	Incremental LYG	Incremental QALYs	ICER versus baseline (£/QALY)	ICER incremental (£/QALY)
Treosulfan	£137,062	8.73	6.40					
Busulfan	£160,821	7.71	5.47	£23,759	-1.01	-0.92	Dominated	Dominated

Abbreviations: ICER, incremental cost-effectiveness ratio; LYG, life years gained; QALYs, quality-adjusted life years

B.3.6.3.5 Scenario 5: Alternative busulfan dosing

An additional scenario analysis was performed assuming a single 3.2mg/kg/day dose for busulfan (rather than 4 x 0.8mg/kg per day doses as per the trial population). Results are shown below in Table 63.

Table 63: Scenario 5: Cost-effectiveness results

Technologies	Total costs (£)	Total LYG	Total QALYs	Incremental costs (£)	Incremental LYG	Incremental QALYs	ICER versus baseline (£/QALY)	ICER incremental (£/QALY)
Treosulfan	£137,082	8.73	6.44					
Busulfan	£159,696	7.71	5.55	£22,614	-1.01	-0.89	Dominated	Dominated

Abbreviations: ICER, incremental cost-effectiveness ratio; LYG, life years gained; QALYs, quality-adjusted life years

B.3.6.3.6 Scenario 6: Separate modelling of AML and MDS

Another scenario was considered where AML and MDS patients were modelled separately in terms of EFS and OS, i.e. using AML and MDS-specific survival data rather than pooling the patients together to generate combined AML + MDS overall survival and event-free survival. To generate pooled results, the overall results for AML and MDS were weighted by the proportion of patients with AML and MDS pre-transplant (63.88% and 36.12% respectively).

The Kaplan-Meier plots and fitted models for each (NMCM lognormal in all instances) are shown below in Figure 28,

Figure 29, Figure 30, Figure 31, Figure 32, Figure 33, Figure 34, and Figure 35.

Figure 28: Kaplan-Meier versus survival model EFS: AML patients, treosulfan arm

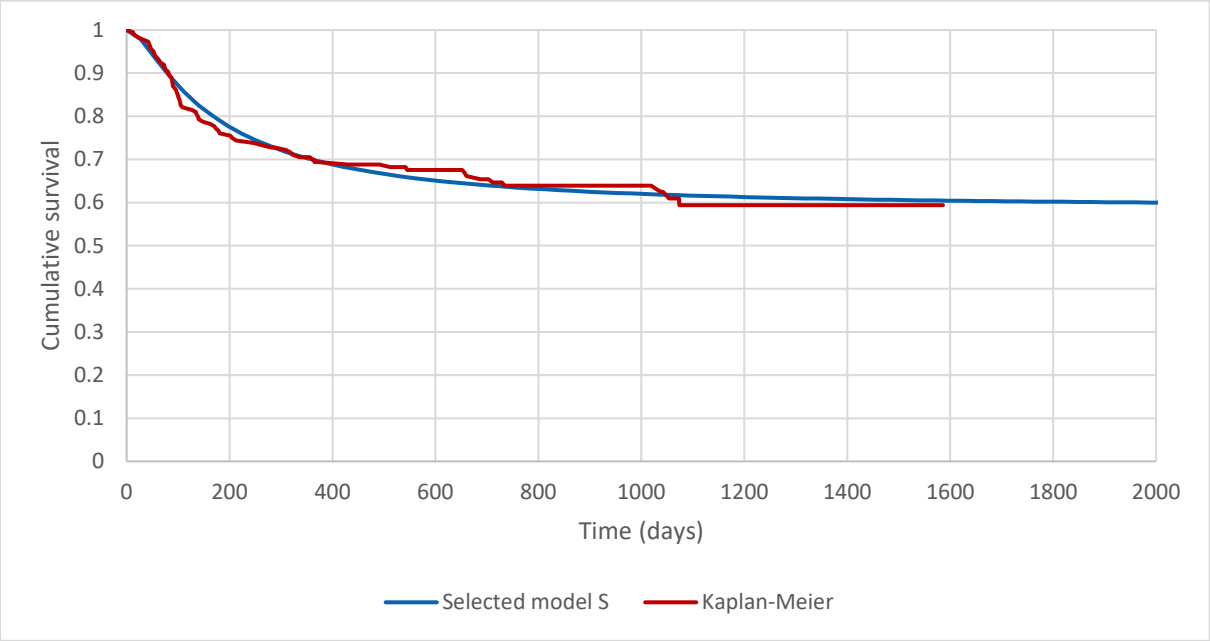


Figure 29: Kaplan-Meier versus survival model EFS: MDS patients, treosulfan arm

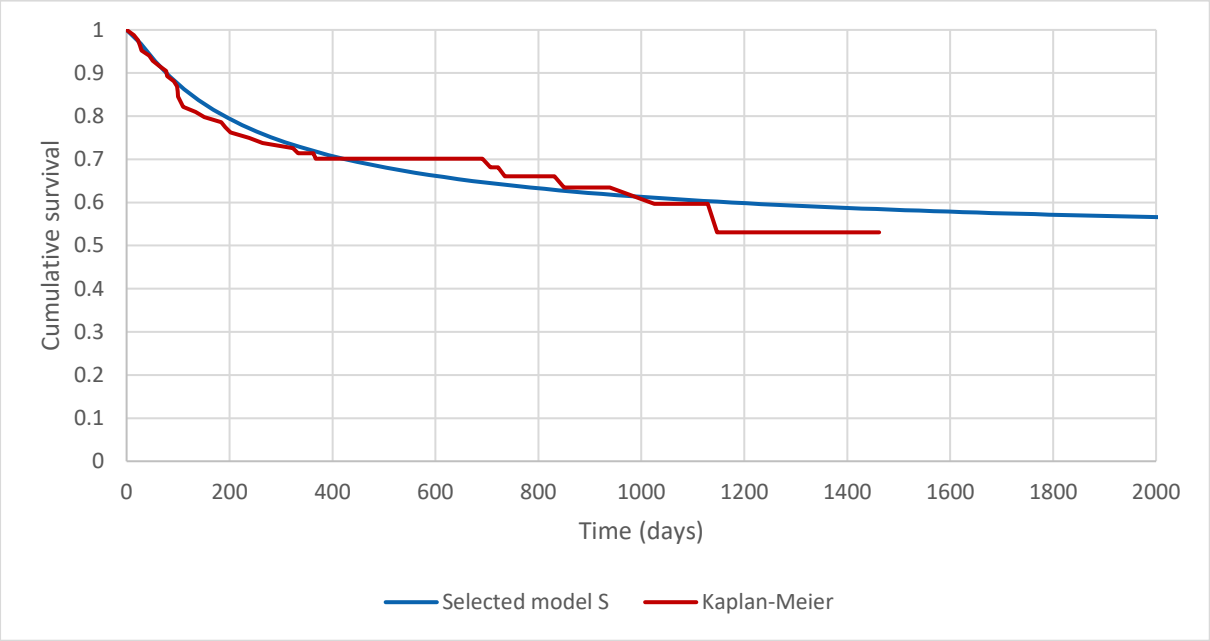


Figure 30: Kaplan-Meier versus survival model EFS: AML patients, busulfan arm

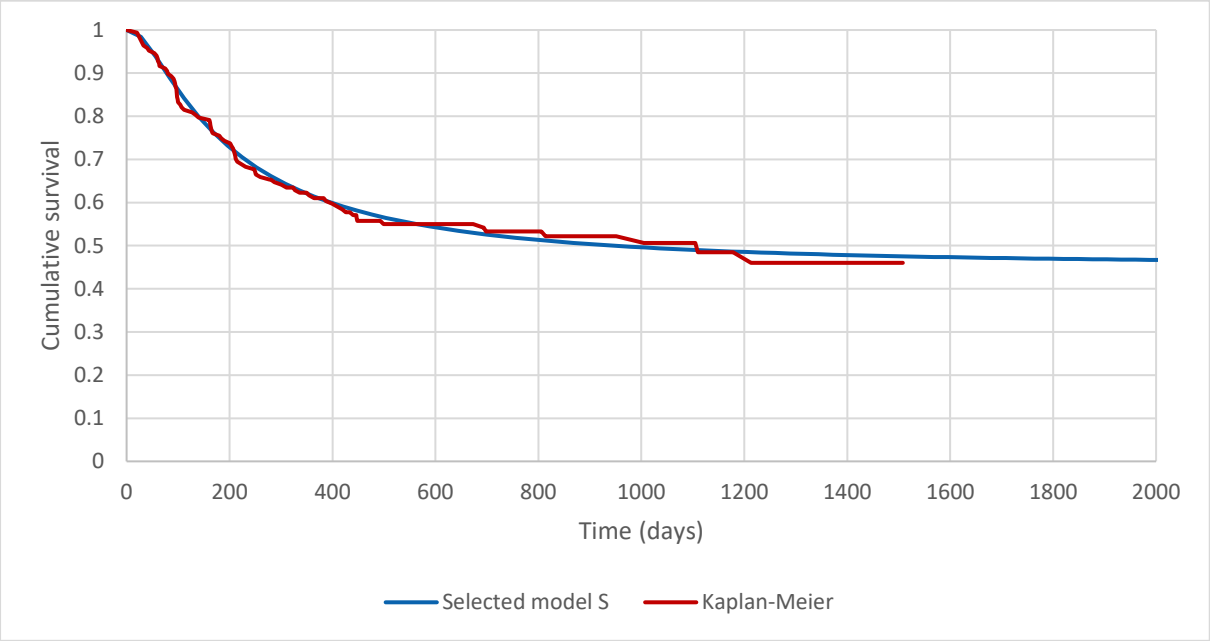


Figure 31: Kaplan-Meier versus survival model EFS: MDS patients, busulfan arm

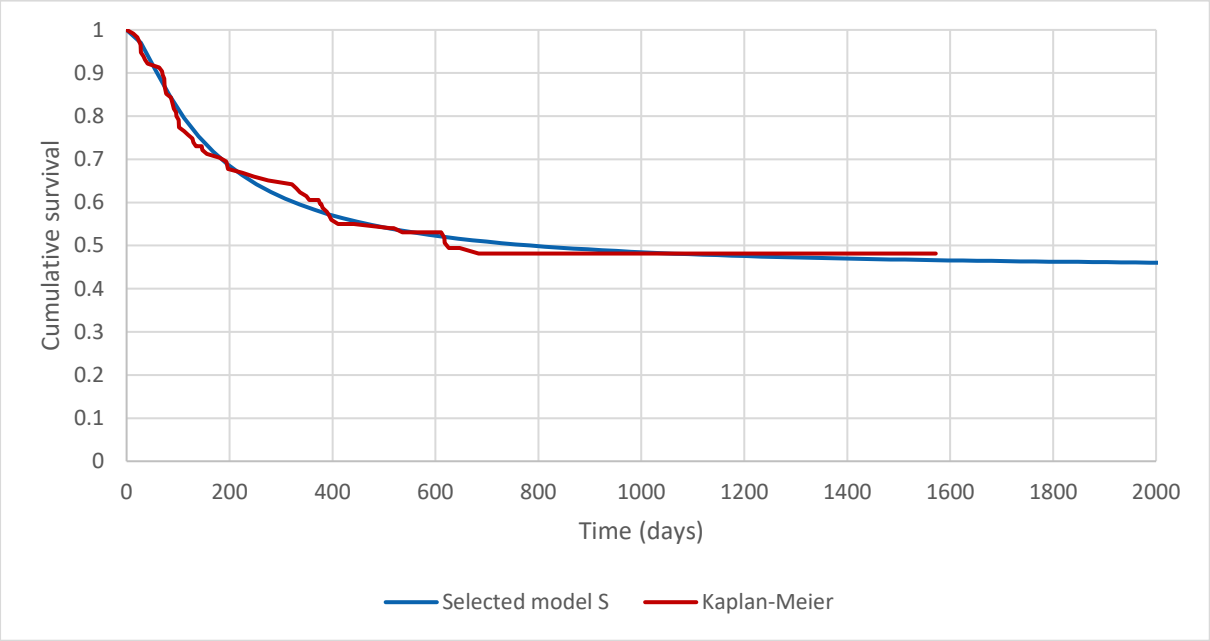


Figure 32: Kaplan-Meier versus survival model OS: AML patients, treosulfan arm

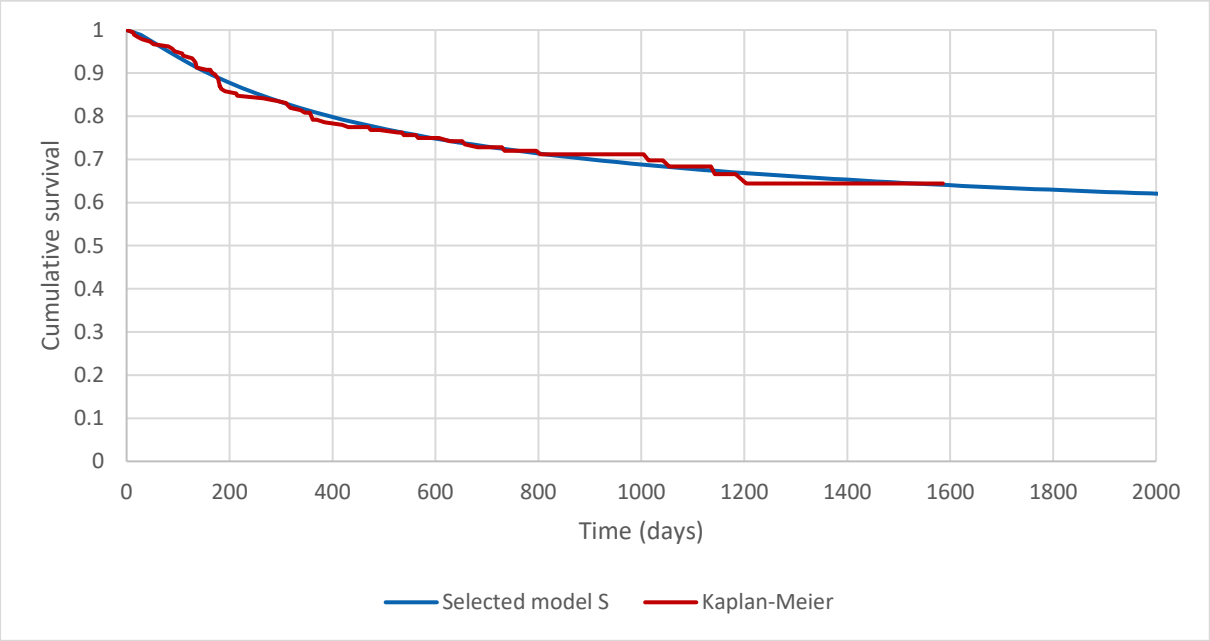


Figure 33: Kaplan-Meier versus survival model OS: MDS patients, treosulfan arm

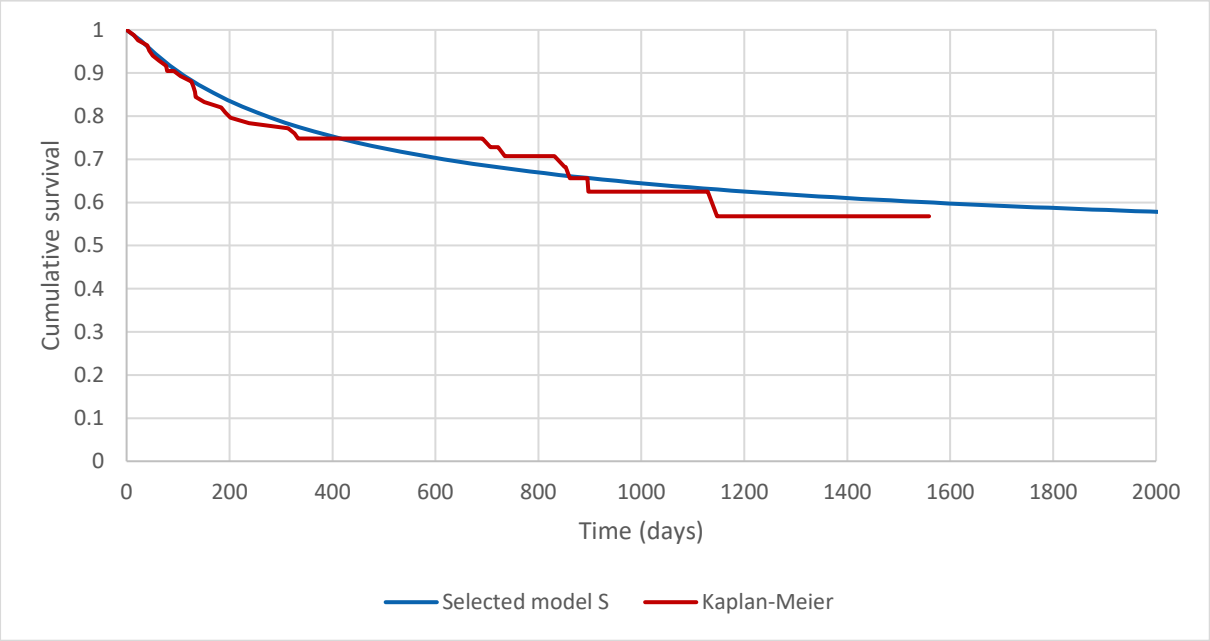


Figure 34: Kaplan-Meier versus survival model OS: AML patients, busulfan arm

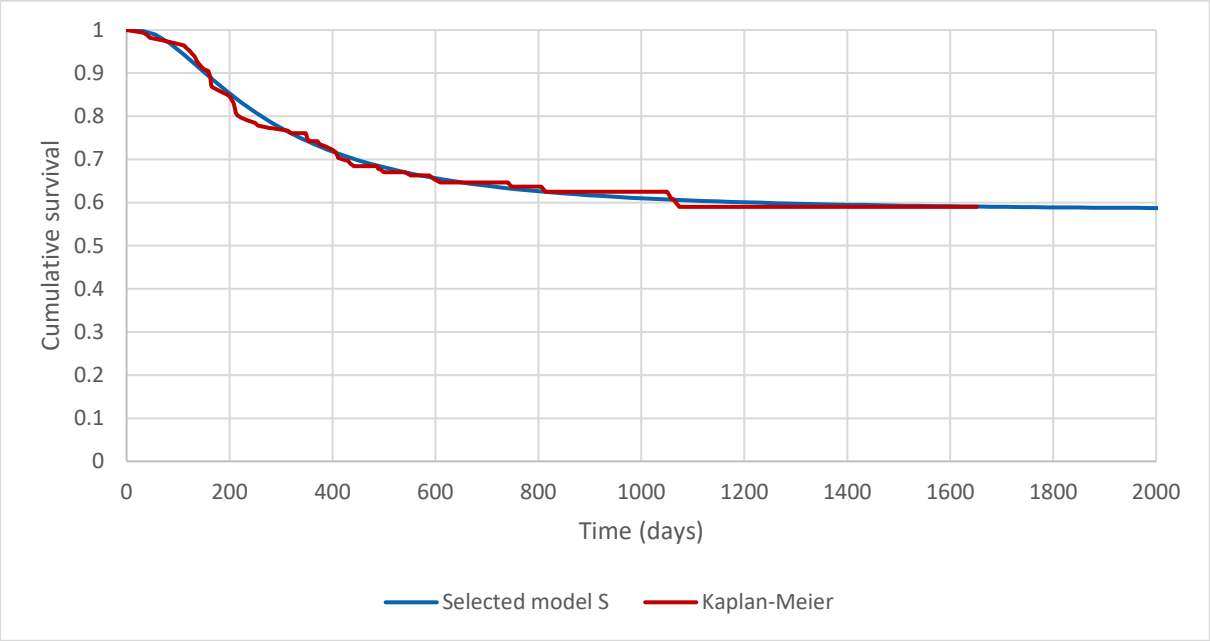
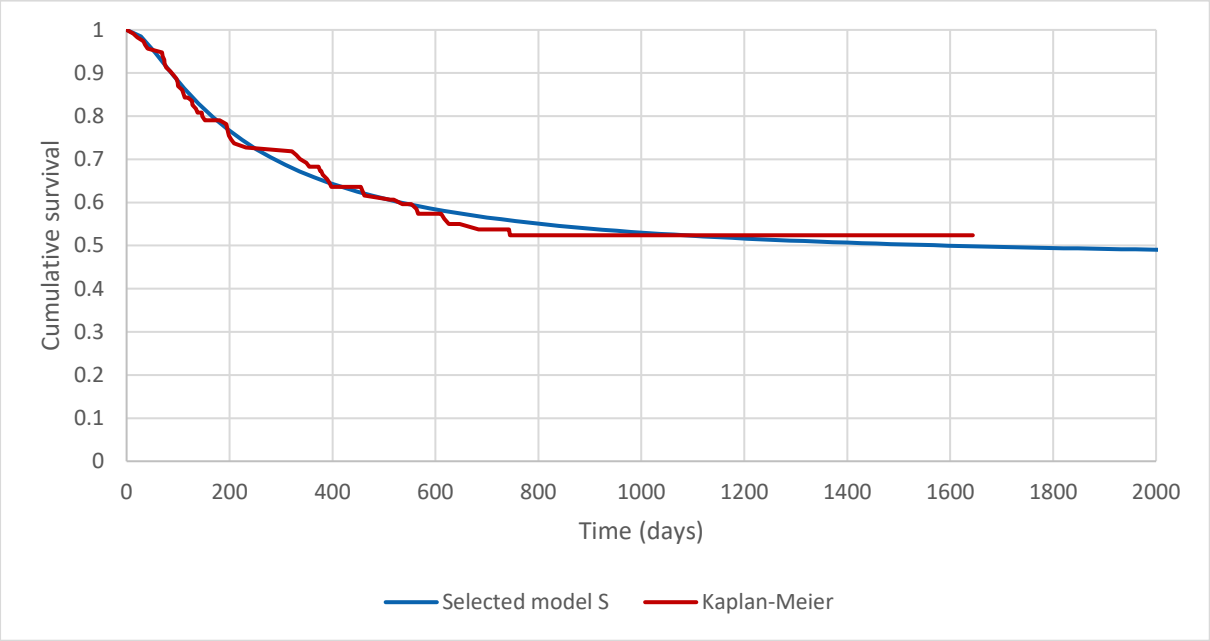


Figure 35: Kaplan-Meier versus survival model OS: MDS patients, busulfan arm



The results of this analysis are shown below in Table 64.

Table 64: Scenario 6: Cost-effectiveness results

Technologies	Total costs (£)	Total LYG	Total QALYs	Incremental costs (£)	Incremental LYG	Incremental QALYs	ICER versus baseline (£/QALY)	ICER incremental (£/QALY)
Treosulfan	£136,970	8.70	6.41					
Busulfan	£162,333	7.73	5.55	£25,363	-0.96	-0.87	Dominated	Dominated

Abbreviations: ICER, incremental cost-effectiveness ratio; LYG, life years gained; QALYs, quality-adjusted life years

B.3.6.3.7 Scenarios 7 and 8: Application of 5- and 10-year time horizons

Further scenarios were considered where 5- and 10-year time horizons were applied. Results are shown below in Table 65 and Table 66 respectively.

Table 65: Scenario 7: Cost-effectiveness results – 5-year time horizon

Technologies	Total costs (£)	Total LYG	Total QALYs	Incremental costs (£)	Incremental LYG	Incremental QALYs	ICER versus baseline (£/QALY)	ICER incremental (£/QALY)
Treosulfan	£116,548	3.32	2.44					
Busulfan	£122,833	2.98	2.15	£6,285	-0.34	-0.29	Dominated	Dominated

Abbreviations: ICER, incremental cost-effectiveness ratio; LYG, life years gained; QALYs, quality-adjusted life-years

Table 66: Scenario 8: Cost-effectiveness results – 10-year time horizon

Technologies	Total costs (£)	Total LYG	Total QALYs	Incremental costs (£)	Incremental LYG	Incremental QALYs	ICER versus baseline (£/QALY)	ICER incremental (£/QALY)
Treosulfan	£124,280	5.57	4.16					
Busulfan	£138,040	4.94	3.61	-£13,761	-0.63	-0.55	Dominated	Dominated

Abbreviations: ICER, incremental cost-effectiveness ratio; LYG, life years gained; QALYs, quality-adjusted life-years

B.3.6.3.8 Scenario 9: Application of gamma survival functions

As described in Section B.3.3.1.3 use of gamma survival functions was explored. Results of the scenario analysis are shown below in Table 67.

Table 67: Scenario 9: Cost-effectiveness results

Technologies	Total costs (£)	Total LYG	Total QALYs	Incremental costs (£)	Incremental LYG	Incremental QALYs	ICER versus baseline (£/QALY)	ICER incremental (£/QALY)
Treosulfan	£143,877	8.67	6.31					
Busulfan	£159,438	7.37	5.23	£15,561	-1.30	-1.09	Dominated	Dominated

Abbreviations: ICER, incremental cost-effectiveness ratio; LYG, life years gained; QALYs, quality-adjusted life-years

B.3.6.4 Summary of sensitivity analyses results

Probabilistic sensitivity analysis indicated that treosulfan was highly cost-effective compared to busulfan, with mean probabilistic incremental costs of -£19,085 and mean probabilistic incremental QALYs of 0.79, generating a mean ICER of -£24,118. Although these estimates differed slightly from the deterministic results (incremental costs of -£23,759, incremental QALYs of 0.89), the CEAC showed that treosulfan was still highly likely to be cost-effective at all thresholds considered, with a probability of cost-effectiveness of 89.9% and 69.6% at WTP thresholds of £30,000 per QALY and £0 per QALY respectively. Increasing the number of simulations did not appear to significantly alter the difference in outcomes observed for the base case and probabilistic results.

A Cholesky decomposition is implemented in the analysis in order to reflect correlation between survival function parameters; however, this is not always sufficient to reduce the skew in the distribution of mean ICERs produced from random draws for the PSA.

Exclusion of the survival inputs from the PSA generated more similar mean probabilistic incremental costs to the base case analysis, albeit with limited impact on the mean probabilistic incremental QALYs. Exclusion of the survival function parameters also generated much smaller variance around costs and QALYs, and higher probabilities of cost-effectiveness for treosulfan at all thresholds considered.

The deterministic sensitivity analysis indicated that the most sensitive inputs in the model were the NMCM lognormal survival function meanlog parameters for treosulfan and busulfan. This was expected given the results from the PSA, and

anticipation that independent variance of correlated parameters would potentially produce extreme values for the ICER.

With the exception of 20% increases for the meanlog coefficients for the treosulfan OS and busulfan EFS curves, which produced ICERs of £34,821 and £33,040 respectively, all parameter variations indicated that treosulfan was dominant over busulfan.

The majority of scenario analyses showed only relatively minor changes in total incremental costs and QALYs. Pooling of individual AML and MDS patient population results produced similar outcomes to the base case model.

As expected, shortening the model time horizon reduced the incremental costs and QALYs. Use of gamma survival functions reduced the difference in total costs by 35%, albeit with an increase of 23% in the difference in total QALYs. Regardless, in all scenario analyses tested, treosulfan dominated busulfan, with lower total costs and higher total QALYs.

B.3.7 Subgroup analysis

Subgroup analysis was performed for AML and MDS patients individually. As in scenario 5, NMCM lognormal functions were applied in for OS and EFS for both treatments and sub-populations.

Results for AML patients are shown in Table 68, with MDS patient results shown in Table 69.

Table 68: AML subgroup cost-effectiveness results

Technologies	Total costs (£)	Total LYG	Total QALYs	Incremental costs (£)	Incremental LYG	Incremental QALYs	ICER versus baseline (£/QALY)	ICER incremental (£/QALY)
Treosulfan	£142,006	8.90	6.54					
Busulfan	£178,633	8.13	5.77	£36,627	-0.76	-0.77	Dominated	Dominated

Abbreviations: ICER, incremental cost-effectiveness ratio; LYG, life years gained; QALYs, quality-adjusted life-years

Table 69: MDS subgroup cost-effectiveness results

Technologies	Total costs (£)	Total LYG	Total QALYs	Incremental costs (£)	Incremental LYG	Incremental QALYs	ICER versus baseline (£/QALY)	ICER incremental (£/QALY)
Treosulfan	£128,063	8.35	6.20					
Busulfan	£133,501	7.03	5.14	£5,172	-1.32	-1.05	Dominated	Dominated

Abbreviations: ICER, incremental cost-effectiveness ratio; LYG, life years gained; QALYs, quality-adjusted life-years

Both subgroup analyses indicated that treosulfan was dominant over busulfan. For the AML subgroup, cost savings for treosulfan were larger than the base case model, with smaller QALY gains. Conversely, the MDS subgroup generated smaller cost savings for treosulfan, albeit with larger QALY gains.

B.3.8 Validation

B.3.8.1 Clinical expert opinion

Two external clinical experts were approached to advise on model parameters. The experts were clinicians experienced in the treatment of AML and MDS with HSCT, and were identified by searching through previous appraisals and committee papers for AML and MDS. A third internal expert from medac was also interviewed with the same questionnaire to provide additional input for the model. All individuals were interviewed by teleconference, and were asked questions primarily related to the following topics:

- Long-term survival assumptions, particularly in relation to SMRs of HSCT patients versus the general population
- Health state utility and cost inputs
- Post-relapse treatment regimens for AML and MDS

The questions submitted to the experts and additional documents for visual representations of the utility functions and long term survival curves are provided in Appendix L.

B.3.8.2 Validation of cost-effectiveness analysis

The model was validated using the Philips' checklist by an independent health economist.[169] A completed version of the checklist is provided in Appendix M1.

Validation of long-term mortality for HSCT survivors, utility inputs and health state costs was performed via consultation with three clinical experts.

B.3.9 Interpretation and conclusions of economic evidence

The results of this economic evaluation demonstrate that the use of treosulfan as a conditioning therapy for HSCT in patients with AML or MDS is highly cost-effective. This is the first economic evaluation directly comparing conditioning regimens for AML and MDS, and therefore the results cannot be compared to previously published analyses.

The economic evaluation is primarily relevant to patients within the Phase III study but may also be relevant to patients with other malignancies which were not covered in this evaluation.

The analysis is relevant to clinical practice in England, although cancer guidelines vary from Trust to Trust, so post-relapse dosing regimens for chemotherapy drugs (and their rate of usage) may vary. Although the majority of patients included in the phase III trial were from a German population, it was anticipated that clinical outcomes for European and English patients would be broadly similar.

The main strength of the analysis is that the partitioned survival structure allows non-linear rates of disease progression to be considered, which would be difficult if an alternative model structure (i.e. Markov) was used. The assumption of a cure point and incorporation of non-mixture-cure and mixture-cure models, and flexibility around long-term mortality assumptions, were also seen as strengths following clinical expert consultation.

A limitation of the analysis was that cost and utility data specific to MDS were generally lacking. This was addressed by following clinical expert feedback and assuming that costs, utilities and long-term mortality would be the same for AML and MDS patients. Another limitation was that treatment sequencing for patients post-relapse was not performed; this was not investigated as subsequent treatments after relapse to AML / MDS were not recorded in the Phase III study.

Uncertainty data were not available for all model parameters, so standard error values had to be assumed using on 20% of the mean value.

Another limitation of the analysis is that the model considers only comparisons of treosulfan and busulfan. However, a network meta-analysis feasibility assessment indicated that comparisons with other therapies (such as cyclophosphamide) were not feasible, due to significant heterogeneity across patient populations and a lack of EFS data in the trials identified.

B.4 References

1. Vertès, A. A. The potential of cytotherapeutics in hematologic reconstitution and in the treatment and prophylaxis of graft-versus-host disease. Chapter I: Current practice and remaining unmet medical needs. *Regen. Med.* **10**, 331–343 (2015).

2. Medac. *Treosulfan Summary of Product Characteristics*. (2018).

3. Medac. Treosulfan Injection for epithelial ovarian cancer Summary of Product Characteristics. *eMC* (2016). Available at: https://www.medicines.org.uk/emc/product/1414/smpc. (Accessed: 5th April 2019)

4. British Society of Blood and Marrow Transplantation (BSBMT). BSBMT Registry 2017. (2017). Available at: http://bsbmt.org/activity/2017/. (Accessed: 5th April 2019)

5. British Society of Blood and Marrow Transplantation (BSBMT). UK Transplant Centre List. Available at: http://bsbmt.org/uk-transplant-centre-list/. (Accessed: 5th April 2019)

6. Abdul Wahid, S. F. *et al.* Comparison of Reduced-Intensity and Myeloablative Conditioning Regimens for Allogeneic Hematopoietic Stem Cell Transplantation in Patients with Acute Myeloid Leukemia and Acute Lymphoblastic Leukemia: A Meta-Analysis. *Stem Cells Dev.* **23**, 2535–2552 (2014).

7. Sorror, M. L. *et al.* Hematopoietic cell transplantation (HCT)-specific comorbidity index: A new tool for risk assessment before allogeneic HCT. *Blood* **106**, 2912–2919 (2005).

8. Hamadani, M., Mohty, M. & Kharfan-Dabaja, M. A. Reduced-intensity conditioning allogeneic hematopoietic cell transplantation in adults with acute myeloid leukemia. *Cancer Control* **18**, 237–245 (2011).

9. Kröger, N., Beelen, D. W. & For the DAG-KBT, D. A. für K. B. e. V.

Conditioning. Allogeneic stem cell transplantation [In German]. *Onkopedia* (2018). Available at: https://www.onkopedia.com/de/onkopedia/guidelines/allgemeine-anforderungen/@@view/html/index.html.

10. Majhail, N. S. *et al.* Indications for Autologous and Allogeneic Hematopoietic Cell Transplantation: Guidelines from the American Society for Blood and Marrow Transplantation. *Biol. Blood Marrow Transplant.* **21**, 1863–1869 (2015).

11. Cornelissen, J. J. & Blaise, D. Hematopoietic stem cell transplantation for patients with AML in first complete remission. *Blood* **127**, 62–70 (2016).

12. Niewerth, D., Creutzig, U., Bierings, M. B. & Kaspers, G. J. L. L. A review on allogeneic stem cell transplantation for newly diagnosed pediatric acute myeloid leukemia. *Blood* **116**, 2205–2214 (2010).

13. Passweg, J. R. *et al.* Is the use of unrelated donor transplantation leveling off in Europe? The 2016 European Society for Blood and Marrow Transplant activity survey report. *Bone Marrow Transplant.* **53**, 1139–1148 (2018).

14. Gratwohl, A. *et al.* Economics and Outcome After Hematopoietic Stem Cell Transplantation: A Retrospective Cohort Study. *EBioMedicine* **2**, 2101–2109 (2015).

15. Passweg, J. R. *et al.* Hematopoietic SCT in Europe: data and trends in 2012 with special consideration of pediatric transplantation. *Bone Marrow Transplant.* **49**, 744–750 (2014).

16. NHS England. 2013/14 NHS Standard Contract for Haematopoietic Stem Cell Transplantation (Adult). 1–34 (2013).

17. McCune, J. S., Bemer, M. J. & Long-Boyle, J. *Pharmacokinetics, Pharmacodynamics and Pharmacogenomics of Immunosuppressants in Allogeneic Haematopoietic Cell Transplantation: Part I*. *Clinical Pharmacokinetics* **55**, (NIH Public Access, 2016).

18. Barba, P. *et al.* Success of an International Learning Health Care System in Hematopoietic Cell Transplantation: The American Society of Blood and Marrow Transplantation Clinical Case Forum. *Biol. Blood Marrow Transplant.* **22**, 564–570 (2016).

19. Deeg, H. J. & Sandmaier, B. M. Who is fit for allogeneic transplantation? *Blood* **116**, 4762–4770 (2010).

20. De Witte, T. *et al.* Allogeneic hematopoietic stem cell transplantation for MDS and CMML: recommendations from an international expert panel. *Blood* **129**, 1753–1762 (2017).

21. European Society for Blood and Marrow Transplantation (EBMT). *The EBMT Handbook Hematopoietic Stem Cell Transplantation and Cellular Therapies*. (2019).

22. Dhawan, R. & Marks, D. I. Who Should Receive a Transplant for Acute Lymphoblastic Leukaemia? *Current Hematologic Malignancy Reports* **12**, 143–152 (2017).

23. Oliansky, D. M. *et al.* The Role of Cytotoxic Therapy with Hematopoietic Stem Cell Transplantation in the Treatment of Adult Acute Lymphoblastic Leukemia: Update of the 2006 Evidence-Based Review. *Biol. Blood Marrow Transplant.* **18**, 18-36.e6 (2012).

24. Gyurkocza, B. & Sandmaier, B. M. Conditioning regimens for hematopoietic cell transplantation: One size does not fit all. *Blood* **124**, 344–353 (2014).

25. Gran, C. *et al.* Treosulfan Conditioning for Allogeneic Transplantation in Multiple Myeloma Improved Overall Survival in Upfront Hematopoietic Stem Cell Transplantation — a Large Retrospective Study By the Chronic Malignancies Working Party of the EBMT. *Blood* **132**, 3464 (2018).

26. Troy, J. D., Atallah, E., Geyer, J. T. & Saber, W. Myelodysplastic syndromes in the United States: An update for clinicians. *Ann. Med.* **46**, 283–289 (2014).

27. Norsworthy, K., Luznik, L. & Gojo, I. New treatment approaches in acute

myeloid leukemia: review of recent clinical studies. *Rev. Recent Clin. Trials* **7**, 224–37 (2012).

28. Grunebaum, E. *et al.* Bone marrow transplantation for severe combined immune deficiency. *J. Am. Med. Assoc.* **295**, 508–518 (2006).

29. Bacigalupo, A. *et al.* Defining the Intensity of Conditioning Regimens: Working Definitions. *Biol. Blood Marrow Transplant.* **15**, 1628–1633 (2009).

30. Champlin, R. Reduced intensity allogeneic hematopoietic transplantation is an established standard of care for treatment of older patients with acute myeloid leukemia. *Best Pract. Res. Clin. Haematol.* **26**, 297–300 (2013).

31. Gratwohl, A. & Carreras, E. Principles of conditioning. in *EBMT - Handbook* 122–137 (2012).

32. Kornblit, B. *et al.* Fludarabine and 2-Gy TBI is superior to 2 Gy TBI as conditioning for HLA-matched related hematopoietic cell transplantation: a phase III randomized trial. *Biol. Blood Marrow Transplant.* **19**, 1340–7 (2013).

33. Hong, S. *et al.* Comparison of non-myeloablative conditioning regimens for lymphoproliferative disorders. *Bone Marrow Transplant.* **50**, 367–374 (2015).

34. Jethava, Y. S. *et al.* Conditioning regimens for allogeneic hematopoietic stem cell transplants in acute myeloid leukemia. *Bone Marrow Transplant.* **52**, 1504–1511 (2017).

35. Finke, J. & Nagler, A. Viewpoint: What is the role of allogeneic haematopoietic cell transplantation in the era of reduced-intensity conditioning - Is there still an upper age limit? A focus on myeloid neoplasia. *Leukemia* **21**, 1357–1362 (2007).

36. Khera, N. *et al.* Nonmalignant late effects and compromised functional status in survivors of hematopoietic cell transplantation. *J. Clin. Oncol.* **30**, 71–77 (2012).

37. Krishnappa, V. *et al.* Acute Kidney Injury in Hematopoietic Stem Cell

Transplantation: A Review. *Int. J. Nephrol.* **2016**, 1–13 (2016).

38. Sakellari, I. *et al.* Survival Advantage and Comparable Toxicity in Reduced-Toxicity Treosulfan-Based versus Reduced-Intensity Busulfan-Based Conditioning Regimen in Myelodysplastic Syndrome and Acute Myeloid Leukemia Patients after Allogeneic Hematopoietic Cell Transplantatio. *Biol. Blood Marrow Transplant.* **23**, 445–451 (2017).

39. Gupta, T., Kannan, S., Dantkale, V. & Laskar, S. Cyclophosphamide plus total body irradiation compared with busulfan plus cyclophosphamide as a conditioning regimen prior to hematopoietic stem cell transplantation in patients with leukemia: a systematic review and meta-analysis. *Hematol. Oncol. Stem Cell Ther.* **4**, 17–29 (2011).

40. Madden, L. M., Ngwube, A. I., Shenoy, S., Druley, T. E. & Hayashi, R. J. Late toxicity of a novel allogeneic stem cell transplant using single fraction total body irradiation for hematologic malignancies in children. *J. Pediatr. Hematol. Oncol.* **37**, e94–e101 (2015).

41. Santos, G. W. *et al.* Marrow Transplantation for Acute Nonlymphocytic Leukemia after Treatment with Busulfan and Cyclophosphamide. *N. Engl. J. Med.* **309**, 1347–1353 (1983).

42. Tuuschka, Tutschka, P., Copelan, E. & Klein, J. Bone marrow transplantation for leukemia following a new busulfan and cyclophosphamide regimen. *Blood* **70**, 1382–1388 (1987).

43. Sengsayadeth, S. *et al.* Reduced intensity conditioning allogeneic hematopoietic cell transplantation for adult acute myeloid leukemia in complete remission -a review from the acute leukemia working party of the EBMT. *Haematologica* **100**, 859–869 (2015).

44. Carella, A. M. Treatment of hematological malignancies with allogeneic nonmyeloablative stem cell transplantation: Conditioning regimens with fludarabine. *Hematology Journal* **5**, S68–S75 (2004).

45. Niederwieser, D. *et al.* Hematopoietic stem cell transplantation activity worldwide in 2012 and a SWOT analysis of the Worldwide Network for Blood and Marrow Transplantation Group including the global survey. *Bone Marrow Transplant.* **51**, 778–785 (2016).

46. Singh, A. K. & McGuirk, J. P. Allogeneic stem cell transplantation: A historical and scientific overview. *Cancer Research* **76**, 6445–6451 (2016).

47. Holtick, U. *et al.* Bone marrow versus peripheral blood allogeneic haematopoietic stem cell transplantation for haematological malignancies in adults. *Cochrane Database of Systematic Reviews* **2014**, (2014).

48. Hilgendorf, I. *et al.* Long-Term Follow-up After Allogeneic Stem Cell Transplantation. *Dtsch. Arztebl. Int.* **112**, 51 (2015).

49. Cooke, K. R. *et al.* The Biology of Chronic Graft-versus-Host Disease: A Task Force Report from the National Institutes of Health Consensus Development Project on Criteria for Clinical Trials in Chronic Graft-versus-Host Disease. *Biology of Blood and Marrow Transplantation* **23**, 211–234 (2017).

50. Mohty, M. *et al.* Sinusoidal obstruction syndrome/veno-occlusive disease: Current situation and perspectives - A position statement from the European Society for Blood and Marrow Transplantation (EBMT). *Bone Marrow Transplantation* **50**, 781–789 (2015).

51. British Society of Blood and Marrow Transplantation (BSBMT), Indications, B., Oct, B. M. T. V. & British Society of Blood and Marrow Transplantation (BSBMT). *BSBMT Indications for BMT Adult*. (2013).

52. Kröger, N. *et al.* Antilymphocyte Globulin for Prevention of Chronic Graft-versus-Host Disease. *N. Engl. J. Med.* **374**, 43–53 (2016).

53. Gratwohl, A. The EBMT risk score. *Bone Marrow Transplant.* **47**, 749–756 (2012).

54. Center for International Blood and Marrow Transplant Research (CIBMTR). Current uses and outcomes of hematopoietic cell Transplantation (HCT). 2018

Summary slides. (2018).

55. NHS England. *Clinical Commissioning Policy: Haematopoietic Stem Cell Transplantation (HSCT) (All Ages): Revised Reference: NHS England B04/P/a*. (2015).

56. Scheulen, M. E. *et al.* Clinical phase I dose escalation and pharmacokinetic study of high-dose chemotherapy with treosulfan and autologous peripheral blood stem cell transplantation in patients with advanced malignancies. *Clin. Cancer Res.* **6**, 4209–4216 (2000).

57. Casper, J. *et al.* Treosulfan and fludarabine: a new toxicity-reduced conditioning regimen for allogeneic hematopoietic stem cell transplantation. *Blood* **103**, 725–731 (2004).

58. Sehouli, J. *et al.* A phase III, open label, randomized multicenter controlled trial of oral versus intravenous treosulfan in heavily pretreated recurrent ovarian cancer: a study of the North-Eastern German Society of Gynecological Oncology (NOGGO). *J. Cancer Res. Clin. Oncol.* **143**, 541–550 (2017).

59. Beelen, D. W. *et al.* Final Results of a Prospective Randomized Multicenter Phase III Trial Comparing Treosulfan/Fludarabine to Reduced Intensity Conditioning with Busulfan/Fludarabine Prior to Allogeneic Hematopoietic Stem Cell Transplantation in Elderly or Comorbid Patients. *Blood* **130**, 521 LP – 521 (2017).

60. Medac. *MC-FludT.14/L Trial II - Clinical Study Report*. (2018).

61. Medac. MC-FludT.6/L. Dose-range Finding Treosulfan-based Conditioning. *ClinicalTrials.gov* (2010). Available at: https://clinicaltrials.gov/ct2/show/NCT01063647. (Accessed: 5th April 2019)

62. Medac. *MC-FludT.6/L - Clinical Study Report*. (2009).

63. Medac. *MC-FludT.7/AML - Clinical Study Report*. (2011).

64. Medac. *MC-FludT.8/MDS - Clinical Study Report*. (2010).

65. Casper, J. *et al.* Allogeneic hematopoietic stem-cell transplantation in patients with hematologic malignancies after dose-escalated treosulfan/fludarabine conditioning. *J. Clin. Oncol.* **28**, 3344–3351 (2010).

66. Casper, J. *et al.* Allogeneic hematopoietic SCT in patients with AML following treosulfan/fludarabine conditioning. *Bone Marrow Transplant.* **47**, 1171–1177 (2012).

67. Ruutu, T. *et al.* Reduced-toxicity conditioning with treosulfan and fludarabine in allogeneic hematopoietic stem cell transplantation for myelodysplastic syndromes:Final results of an international prospective phase II trial. *Haematologica* **96**, 1344–1350 (2011).

68. Medac. MC-FludT.16/NM. Treosulfan-based Versus Busulfan-based Conditioning in Paediatric Patients With Non-malignant Diseases. *ClinicalTrials.gov* (2018). Available at: https://clinicaltrials.gov/ct2/show/NCT02349906. (Accessed: 8th April 2019)

69. European Society for Blood and Marrow Transplantation (EBMT), Iacobelli, S., Koster, L. & Biezen, A. Van. *Re-analysis of EBMT-registry data on Fludarabine/Melphalan and Busulfan/Cyclophosphamide based conditioning treatment compared to Fludarabine/Treosulfan based conditioning of adult AML and MDS patients treated in MC-FludT.14/L phase III by matched pairs.* . (2019).

70. Gorin, N. C. *et al.* Results of genoidentical hemopoietic stem cell transplantation with reduced intensity conditioning for acute myelocytic leukemia: Higher doses of stem cells infused benefit patients receiving transplants in second remission or beyond - The Acute Leukemia . *J. Clin. Oncol.* **24**, 3959–3966 (2006).

71. Kroger, N. *et al.* Allogeneic stem cell transplantation after a fludarabine/busulfan-based reduced-intensity conditioning in patients with myelodysplastic syndrome or secondary acute myeloid leukemia. *Ann. Hematol.* **82**, 336–342 (2003).

72. Medac. *MC-FludT.14/L Trial II Interim Clinical Study Report - 476 patients. To compare Treosulfan-based conditioning therapy with Busulfan-based reduced-intensity conditioning (RIC) prior to alloHSCT patients with AML or MDS considered ineligible to standard con*. (2017).

73. Swerdlow, S. H. *et al. WHO Classification of Tumours: Pathology and Genetics. Tumours of Haematopoietic and Lymphoid Tissues*. *IARC Press, Lyon 2008.* (2008). doi:10.1046/j.1365-2257.2003.00518.x

74. National Health Service (NHS). Stem cell and bone marrow transplants. *www.nhs.uk* (2018). Available at: https://www.nhs.uk/conditions/stem-cell-transplant/#. (Accessed: 5th April 2019)

75. DuBois, D. & DuBois, E. A formula to estimate the approximate surface area if height and weight be known. *Arch Intern Med* **17**, 863–87 (1916).

76. Medac. *MC FludT.14/L Trial II - Statistical Analysis Plan*. (2018).

77. McCurdy, S. R. *et al.* Comparable composite endpoints after HLA-matched and HLA-haploidentical transplantation with post-transplantation cyclophosphamide. *Haematologica* **102**, 391–400 (2017).

78. Potter, V. T. *et al.* Long-Term Outcomes of Alemtuzumab-Based Reduced-Intensity Conditioned Hematopoietic Stem Cell Transplantation for Myelodysplastic Syndrome and Acute Myelogenous Leukemia Secondary to Myelodysplastic Syndrome. *Biol. Blood Marrow Transplant.* **20**, 111–117 (2014).

79. Russell, N. H. *et al.* A comparative assessment of the curative potential of reduced intensity allografts in acute myeloid leukaemia. *Leukemia* **29**, 1478–1484 (2015).

80. Killick, S. B. *et al.* Guidelines for the diagnosis and management of adult myelodysplastic syndromes. *Br. J. Haematol.* **164**, 503–525 (2014).

81. National Institute Health Research (NIHR). AML15: Protcol for patients aged under 60. (2004).

82. National Institute Health Research (NIHR). AML16: A trial for older patients with acute AML and high risk myelodysplastic syndrome. 1–46 (2006).

83. National Institute for Health Research (NIHR). AML18: Risk Myelodysplastic Syndrome - Protocol under development. (2019). Available at: https://ukctg.nihr.ac.uk/trials/trial-details/trial-details?trialNumber=ISRCTN78449203.

84. Yorkshire and Humberside Clinical Networks. *Clinical Guidelines for Leukaemia and other Myeloid Disorders – Myeloproliferative Neoplasms*. (2017).

85. Beelen, D. *et al.* Final Evaluation of a Clinical Phase III Trial Comparing Treosulfan to Busulfan-Based Conditioning Therapy Prior to Allogeneic Hematopoietic Stem Cell Transplantation of Adult Acute Myeloid Leukemia and Myelodysplastic Syndrome Patients Ineligible to Stan. *Biol. Blood Marrow Transplant.* **25**, S3 (2019).

86. Kroeger, N. *et al.* Reduced Intensity Vs. Myeloablative Conditioning Followed By Allogeneic Stem Cell Transplantation for Patients with Myelodysplastic Syndrome: Long Term Follow-up of a Prospective Randomized EBMT Phase III Study (RICMAC-Trial). *Blood* **132**, 1019 (2018).

87. Kröger, N. *et al.* Dose-Reduced Versus Standard Conditioning Followed by Allogeneic Stem-Cell Transplantation for Patients With Myelodysplastic Syndrome: A Prospective Randomized Phase III Study of the EBMT (RICMAC Trial). *J Clin Oncol* **35**, 2157–2164 (2017).

88. Liu, H. *et al.* Busulfan plus fludarabine as a myeloablative conditioning regimen compared with busulfan plus cyclophosphamide for acute myeloid leukemia in first complete remission undergoing allogeneic hematopoietic stem cell transplantation: A prospective and multicen. *J. Hematol. Oncol.* (2013). doi:10.1186/1756-8722-6-15

89. Rambaldi, A. *et al.* Busulfan plus cyclophosphamide versus busulfan plus fludarabine as a preparative regimen for allogeneic haemopoietic stem-cell

transplantation in patients with acute myeloid leukaemia: An open-label, multicentre, randomised, phase 3 trial. *Lancet Oncol.* **16**, 1525–1536 (2015).

90. Kröger, N. *et al.* Reduced Intensity Vs. Myeloablative Conditioning Followed By Allogeneic Stem Cell Transplantation for Patients with Myelodysplastic Syndrome: Long Term Follow-up of a Prospective Randomized EBMT Phase III Study (RICMAC-Trial). *Blood* **132**, 1019 (2018).

91. National Institute for Health and Care Excellence (NICE). TA523: Midostaurin for untreated acute myeloid leukaemia. *Technology Appraisal Guidance* (2018). Available at: https://www.nice.org.uk/guidance/ta523. (Accessed: 17th April 2019)

92. National Institute for Health and Care Excellence (NICE). TA552: Liposomal cytarabine–daunorubicin for untreated acute myeloid leukaemia. *Technology Appraisal Guidance* (2018). Available at: https://www.nice.org.uk/guidance/ta552. (Accessed: 17th April 2019)

93. Le Corroller S, Anne-Gaelle, C Siani, R Tabrizi, M Michallet, J-O Bay, J-M Boher, C Faucher, D. B. The choice of effectiveness criteria affects conclusions of economic evaluation of newer allogeneic bone marrow transplantation modalities :example based on a randomised multicenter trial comparing two reduced intensity conditioning regimen (FLU-BU-ATG) v. *Bone Marrow Transplant.* **52**, S312–S313 (2017).

94. Batty, N., Yin, Y. & Wetzler, M. Decitabine is more cost effective than cytarabine and daunorubicin in elderly acute myeloid leukemia patients. *J. Cancer Res. Ther.* **2**, 68–73 (2014).

95. Goss, T. F. *et al.* Cost effectiveness of lenalidomide in the treatment of transfusion- dependent myelodysplastic syndromes in the United States. *Cancer Control* **13**, 17–25 (2006).

96. Levy, A. R. *et al.* Cost-effectiveness in Canada of azacitidine for the treatment of higher-risk myelodysplastic syndromes. *Curr. Oncol.* **21**, 29 (2014).

97. Pan, F. *et al.* Economic Analysis of Decitabine Versus Best Supportive Care in the Treatment of Intermediate- and High-Risk Myelodysplastic Syndromes From a US Payer Perspective. *Clin. Ther.* **32**, 2444–2456 (2010).

98. Stein, E. *et al.* Cost Effectiveness of Midostaurin in the Treatment of Newly Diagnosed FLT3-Mutated Acute Myeloid Leukemia in the United States. *Pharmacoeconomics* **37**, 239–253 (2019).

99. Tremblay, G., Dolph, M., Patel, S., Brandt, P. & Forsythe, A. Cost-effectiveness analysis for midostaurin versus standard of care in acute myeloid leukemia in the United Kingdom. *Cost Eff. Resour. Alloc.* **16**, 33 (2018).

100. Uyl-de Groot, C. A. *et al.* Cost-effectiveness and quality-of-life assessment of GM-CSF as an adjunct to intensive remission induction chemotherapy in elderly patients with acute myeloid leukemia. *Br. J. Haematol.* **100**, 629–36 (1998).

101. Crespo, C., Moreno, E., Sierra, J., Serip, S. & Rubio, M. Cost-effectiveness analysis of azacitidine in the treatment of high-risk myelodysplastic syndromes in Spain. *Health Econ. Rev.* **3**, 1–10 (2013).

102. National Institute for Health and Care Excellence (NICE). TA399: Azacitidine for treating acute myeloid leukaemia with more than 30% bone marrow blasts. *Technology Appraisal Guidance* (2016). Available at: https://www.nice.org.uk/guidance/ta399. (Accessed: 17th April 2019)

103. National Institute for Health and Care Excellence (NICE). TA545: Gemtuzumab ozogamicin for untreated acute myeloid leukaemia. *Technology Appraisal Guidance* (2018). Available at: https://www.nice.org.uk/guidance/ta545. (Accessed: 17th April 2019)

104. NHS Improvement. National schedule of reference costs 2017/18. Available at: https://improvement.nhs.uk/resources/reference-costs/. (Accessed: 1st April 2019)

105. National Institute for Health and Care Excellence (NICE). TA218: Azacitidine

for the treatment of myelodysplastic syndromes, chronic myelomonocytic leukaemia and acute myeloid leukaemia. *Technology Appraisal Guidance* (2011). Available at: https://www.nice.org.uk/guidance/ta218. (Accessed: 17th April 2019)

106. National Institute for Health and Care Excellence (NICE). TA322: Lenalidomide for treating myelodysplastic syndromes associated with an isolated deletion 5q cytogenetic abnormality. *Technology Appraisal Guidance* (2014). Available at: https://www.nice.org.uk/guidance/ta322. (Accessed: 17th April 2019)

107. National Institute for Health and Care Excellence Decision Support Unit (NICE DSU). TSD 14: Survival analysis for economic evaluations alongside clinical trials – extrapolation with patient-level data. doi:10.1177/0272989X13497998

108. Marks, D. I. *et al.* The outcome of full-intensity and reduced-intensity conditioning matched sibling or unrelated donor transplantation in adults with Philadelphia chromosome-negative acute lymphoblastic leukemia in first and second complete remission. *Blood* **116**, 366–374 (2010).

109. Martin, P. J. *et al.* Life expectancy in patients surviving more than 5 years after hematopoietic cell transplantation. *J. Clin. Oncol.* **28**, 1011–1016 (2010).

110. Kurosawa, S. *et al.* A Markov decision analysis of allogeneic hematopoietic cell transplantation versus chemotherapy in patients with acute myeloid leukemia in first remission. *Blood* **117**, 2113–2120 (2011).

111. Aristides, M., Barlev, A., Barber, B., Gijsen, M. & Quinn, C. Population preference values for health states in relapsed or refractory B-precursor acute lymphoblastic leukemia in the United Kingdom. *Health Qual. Life Outcomes* **13**, 181 (2015).

112. Stauder, R. *et al.* Health-Related Quality of Life is Substantially Impaired in Lower-Risk MDS when Compared with Reference Populations and Significantly Affects Overall Survival. *Leuk. Res.* **55**, S12 (2017).

113. Stauder, R. *et al.* Health-related quality of life in lower-risk MDS patients

compared with age- and sex-matched reference populations: a European LeukemiaNet study. *Leukemia* **32**, 1380–1392 (2018).

114. Castejón, N. *et al.* Social preferences for health states associated with acute myeloid leukemia for patients undergoing treatment in the United Kingdom. *Health Qual. Life Outcomes* **16**, 66 (2018).

115. Joshi, N. *et al.* Health State Utilities for Acute Myeloid Leukaemia: A Time Trade-off Study. *Pharmacoeconomics* **37**, 85–92 (2019).

116. Leunis, A., Redekop, W. K., Uyl-de Groot, C. A. & Löwenberg, B. Impaired health-related quality of life in acute myeloid leukemia survivors: A single-center study. *Eur. J. Haematol.* **93**, 198–206 (2014).

117. Mamolo, C. M. *et al.* Cross-Sectional Survey of Symptoms and Health-Related Quality of Life of Adults with De Novo Acute Myeloid Leukemia (AML) in Clinical Practice. *Blood* **130**, 5660 LP – 5660 (2017).

118. Stein, E. M. *et al.* Assessing utility values for treatment-related health states of acute myeloid leukemia in the United States. *Health Qual. Life Outcomes* **16**, 193 (2018).

119. Kurosawa, S. *et al.* Decision Analysis of Postremission Therapy in Cytogenetically Intermediate-Risk Acute Myeloid Leukemia: The Impact of FLT3 Internal Tandem Duplication, Nucleophosmin, and CCAAT/Enhancer Binding Protein Alpha. *Biol. Blood Marrow Transplant.* **22**, 1125–1132 (2016).

120. Szende, A. *et al.* Valuation of transfusion-free living in MDS: results of health utility interviews with patients. *Health Qual. Life Outcomes* **7**, 81 (2009).

121. Kayastha, N. *et al.* The impact of remission status on patients' experiences with acute myeloid leukemia (AML): an exploratory analysis of longitudinal patient-reported outcomes data. *Support. Care Cancer* **26**, 1437–1445 (2018).

122. Buckley, S. A., Jimenez-Sahagun, D., Othus, M., Walter, R. B. & Lee, S. J. Quality of life from the perspective of the patient with acute myeloid leukemia. *Cancer* **124**, 145–152 (2018).

123. Timilshina, N., Breunis, H., Tomlinson, G., Brandwein, J. & Alibhai, S. M. H. H. Do quality of life, physical function, or the Wheatley index at diagnosis predict 1-year mortality with intensive chemotherapy in older acute myeloid leukemia patients? *Leuk. Res.* **47**, 142–148 (2016).

124. Ramos, F. *et al.* Impact of anaemia on health-related quality of life and cardiac remodelling in patients with lower risk myelodysplastic syndromes. Results of GlobQoL study. *Eur. J. Cancer Care (Engl).* **26**, e12426 (2017).

125. Messerer, D. *et al.* Impact of different post-remission strategies on quality of life in patients with acute myeloid leukemia. *Haematologica* **93**, 826–833 (2008).

126. McKenzie, L. & Van Der Pol, M. Mapping the EORTC QLQ C-30 onto the EQ-5D instrument: The potential to estimate QALYs without generic preference data. *Value Heal.* **12**, 167–171 (2009).

127. Proskorovsky, I. *et al.* Mapping EORTC QLQ-C30 and QLQ-MY20 to EQ-5D in patients with multiple myeloma. *Health Qual. Life Outcomes* **12**, 35 (2014).

128. Kurosawa, S. *et al.* Patient-reported quality of life after allogeneic hematopoietic cell transplantation or chemotherapy for acute leukemia. *Bone Marrow Transplant.* **50**, 1241–1249 (2015).

129. Kurosawa, S. *et al.* Decision Analysis of Allogeneic Hematopoietic Stem Cell Transplantation Versus Chemotherapy in Cytogenetically Standard-Risk Acute Myeloid Leukemia in First Complete Remission: The Impact of FLT3-ITD Profile. *Blood* **124**, (2014).

130. Slovacek, L., Slovackova, B., Jebavy, L. & Macingova, Z. Psychosocial, health and demographic characteristics of quality of life among patients with acute myeloid leukemia and malignant lymphoma who underwent autologous hematopoietic stem cell transplantation. *Sao Paulo Med. J.* **125**, 359–361 (2007).

131. Grulke, N., Albani, C. & Bailer, H. Quality of life in patients before and after

haematopoietic stem cell transplantation measured with the European Organization for Research and Treatment of Cancer (EORTC) Quality of Life Core Questionnaire QLQ-C30. *Bone Marrow Transplantation* **47**, 473–482 (2012).

132. Perić, Z. *et al.* Which questionnaires should we use to evaluate quality of life in patients with chronic graft-vs-host disease? *Croat. Med. J.* **57**, 6–15 (2016).

133. Lee, S. J. *et al.* Development and validation of a scale to measure symptoms of chronic graft-versus-host disease. *Biol. Blood Marrow Transplant.* **8**, 444–452 (2002).

134. Lee, S. J. *et al.* Quality of life associated with acute and chronic graft-versus-host disease. *Bone Marrow Transplant.* **38**, 305–310 (2006).

135. Dakin, H., Abel, L., Burns, R. & Yang, Y. Review and critical appraisal of studies mapping from quality of life or clinical measures to EQ-5D: An online database and application of the MAPS statement. *Health Qual. Life Outcomes* **16**, 31 (2018).

136. Cheung, Y. B. *et al.* Mapping the English and Chinese versions of the functional assessment of cancer therapy-general to the EQ-5D utility index. *Value Heal.* **12**, 371–376 (2009).

137. Khan, I. & Morris, S. A non-linear beta-binomial regression model for mapping EORTC QLQ- C30 to the EQ-5D-3L in lung cancer patients: A comparison with existing approaches. *Health Qual. Life Outcomes* **12**, (2014).

138. Crott, R. & Briggs, A. Mapping the QLQ-C30 quality of life cancer questionnaire to EQ-5D patient preferences. *Eur. J. Heal. Econ.* **11**, 427–434 (2010).

139. Kontodimopoulos, N., Aletras, V. H., Paliouras, D. & Niakas, D. Mapping the cancer-specific EORTC QLQ-C30 to the preference-based EQ-5D, SF-6D, and 15D instruments. *Value Heal.* **12**, 1151–1157 (2009).

140. Marriott, E. R., van Hazel, G., Gibbs, P. & Hatswell, A. J. Mapping EORTC-

QLQ-C30 to EQ-5D-3L in patients with colorectal cancer. *J. Med. Econ.* **20**, 193–199 (2016).

141. Hill, G. Algorithm 396: Student's t-Quantiles. *Commun. ACM* **13**, 619–620 (1970).

142. Longworth, L. *et al.* Use of generic and condition-specific measures of health-related quality of life in NICE decision-making: A systematic review, statistical modelling and survey. *Health Technol. Assess. (Rockv).* **18**, 1–224 (2014).

143. Doble, B. & Lorgelly, P. Mapping the EORTC QLQ-C30 onto the EQ-5D-3L: assessing the external validity of existing mapping algorithms. *Qual. Life Res.* **25**, 891–911 (2016).

144. Rowen, D., Young, T., Brazier, J. & Gaugris, S. Comparison of generic, condition-specific, and mapped health state utility values for multiple myeloma cancer. *Value Health* **15**, 1059–1068 (2012).

145. Hiramoto, N. *et al.* Positive impact of chronic graft-versus-host disease on the outcome of patients with de novo myelodysplastic syndrome after allogeneic hematopoietic cell transplantation: A single-center analysis of 115 patients. *Eur. J. Haematol.* **92**, 137–146 (2014).

146. Maund, E. *et al.* Management of frozen shoulder: A systematic review and cost-effectiveness analysis. *Health Technology Assessment* **16**, (2012).

147. Longo M, Cohen D, Hood K, R. M. *Deriving an 'enhanced' EuroQoL from SF-36. Presented at the Health Economics Study Group (HESG) meeting, July 2000, Nottingham.* (2000).

148. Longworth, L. *Estimating quality adjusted life years where health-related utility data are missing. PhD thesis. Brunel University.* (2007).

149. Rowen, D., Brazier, J. & Roberts, J. Mapping SF-36 onto the EQ-5D index: How reliable is the relationship? *Health Qual. Life Outcomes* **7**, 27 (2009).

150. Ara, R. & Brazier, J. Deriving an algorithm to convert the eight mean SF-36

dimension scores into a mean EQ-5D preference-based score from published studies (where patient level data are not available). *Value Heal.* **11**, 1131–1143 (2008).

151. Rivero-Arias, O., Gray, A. & Ramos-Goñi, J. Estimating the association between SF-12 responses to EQ5-D utility values using a reponse mapping and a parameter uncertainty approach. *Value Heal.* **13**, A248 (2010).

152. Franks, P., Lubetkin, E. I., Gold, M. R. & Tancredi, D. J. Mapping the SF-12 to Preference-Based Instruments. *Med. Care* **41**, 1277–1283 (2003).

153. Lawrence, W. F. & Fleishman, J. A. Predicting EuroQoL EQ-5D Preference Scores from the SF-12 Health Survey in a Nationally Representative Sample. *Med. Decis. Mak.* **24**, 160–169 (2004).

154. Franks, P., Lubetkin, E. I., Gold, M. R., Tancredi, D. J. & Jia, H. Mapping the SF-12 to the EuroQol EQ-5D index in a national US sample. *Medical Decision Making* **24**, 247–254 (2004).

155. Pickard, A. S., Wang, Z., Walton, S. M. & Lee, T. A. Are decisions using cost-utility analyses robust to choice of SF-36/SF-12 preference-based algorithm? *Health Qual. Life Outcomes* **3**, 11 (2005).

156. Snedecor, S. J. *et al.* Cost-effectiveness of eszopiclone for the treatment of adults with primary chronic insomnia. *Sleep* **32**, 817–24 (2009).

157. Shulman, H. M. *et al.* Chronic graft-versus-host syndrome in man. A long-term clinicopathologic study of 20 Seattle patients. *Am. J. Med.* **69**, 204–17 (1980).

158. Filipovich, A. H. *et al.* National Institutes of Health consensus development project on criteria for clinical trials in chronic graft-versus-host disease: I. Diagnosis and staging working group report. *Biol. Blood Marrow Transplant.* **11**, 945–56 (2005).

159. Sutton, L. *et al.* Autologous stem cell transplantation as a first-line treatment strategy for chronic lymphocytic leukemia: a multicenter, randomized, controlled trial from the SFGM-TC and GFLLC for the Société Française de

Greffe de Moelle et de Thé rapie Cellulaire (SFGM. *Blood* **117**, 6109–19 (2011).

160. Ara, R. & Brazier, J. E. Populating an economic model with health state utility values: Moving toward better practice. *Value Heal.* **13**, 509–518 (2010).

161. Joint Formulary Committee. British National Formulary (on line). *BMJ Group and Pharmaceutical Press* (2019). Available at: https://www.medicinescomplete.com/. (Accessed: 1st April 2019)

162. Department of Health and Social Care. NHS reference costs 2014 to 2015 - GOV.UK. (2015). Available at: https://www.gov.uk/government/publications/nhs-reference-costs-2014-to-2015. (Accessed: 1st April 2019)

163. Kloos, R. Q. H. *et al.* A cost analysis of individualized asparaginase treatment in pediatric acute lymphoblastic leukemia. *Pediatr. Blood Cancer* **64**, (2017).

164. Espérou, H. *et al.* Predicting the costs of allogeneic sibling stem-cell transplantation: results from a prospective, multicenter, French study. *Transplantation* **77**, 1854–8 (2004).

165. Khera, N. *et al.* Costs of Allogeneic Hematopoietic Cell Transplantation Using Reduced Intensity Conditioning Regimens. *Oncologist* **19**, 639–644 (2014).

166. Organisation for Economic Co-operation and Development (OECD). Conversion Rates - Purchasing Power Parities (PPP). (2008). Available at: https://data.oecd.org/conversion/purchasing-power-parities-ppp.htm. (Accessed: 1st April 2019)

167. Curtis, L. & Burns, A. Unit Costs of Health and Social Care 2018. Personal Social Services Research Unit, University of Kent, Canterbury. (2018). Available at: https://www.pssru.ac.uk/project-pages/unit-costs/unit-costs-2018. (Accessed: 1st April 2019)

168. NHS England & Monitor. NHS National Tariff Payment System 2016/17 - GOV.UK. (2017). Available at:

https://www.gov.uk/government/publications/nhs-national-tariff-payment-system-201617. (Accessed: 1st April 2019)

169. Philips, Z. *et al.* Review of guidelines for good practice in decision-analytic modelling in health technology assessment. *Health Technology Assessment* **8**, (2004).

NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE

Single technology appraisal

Treosulfan with fludarabine for malignant disease before allogeneic stem cell transplant [ID1508]

Clarification questions

June 2019

File name	Version	Contains confidential information	Date
ID1508 treosulfan – ERG Clarification Response	FINAL	Yes (AIC)	1 July 2019

Section A: Literature searching

A1. Priority question: Please justify why the comparator, thiotepa, as per the final scope, has not been searched for in any of the database or congress searches.

Please consider re-running all searches as well as screening with this comparator included to ensure nothing has been missed.

Treosulfan and busulfan are stem cell depleting agents; however, thiotepa does not enhance stem cell depletion. The primary treatment effect of thiotepa within a conditioning regimen is to provide additional immunosuppression thereby preventing graft rejection.[1] Therefore, thiotepa is always combined with a stem cell toxic treatment modality (e.g. treosulfan, busulfan, melphalan or total body irradiation). Therefore thiotepa was effectively included in the search, as an adjunct to the primary cell depleting agents included in the search terms.

Today, thiotepa is usually combined with treosulfan or busulfan, especially in paediatric patients. In medac's trial MC-FludT.17/M, on investigators discretion, for additional immune-suppression, subjects could receive thiotepa in 2 single doses of 5 mg/kg given on Day -2. Indeed, 65 out of 70 subjects received such additional treatment. Furthermore, medac is currently running a trial comparing conditioning with treosulfan with fludarabine ± thiotepa versus busulfan with fludarabine ± thiotepa in paediatric patients with non-malignant diseases. In this trial, the additional thiotepa can be administered as investigator's choice for a given patient (MC-FludT.16/NM) and again the majority of patients treated so far received additional thiotepa in both groups.

Thiotepa cannot therefore be considered as an alternative or comparator to treosulfan but as an important adjunct to the treosulfan with fludarabine or busulfan with fludarabine regimen.

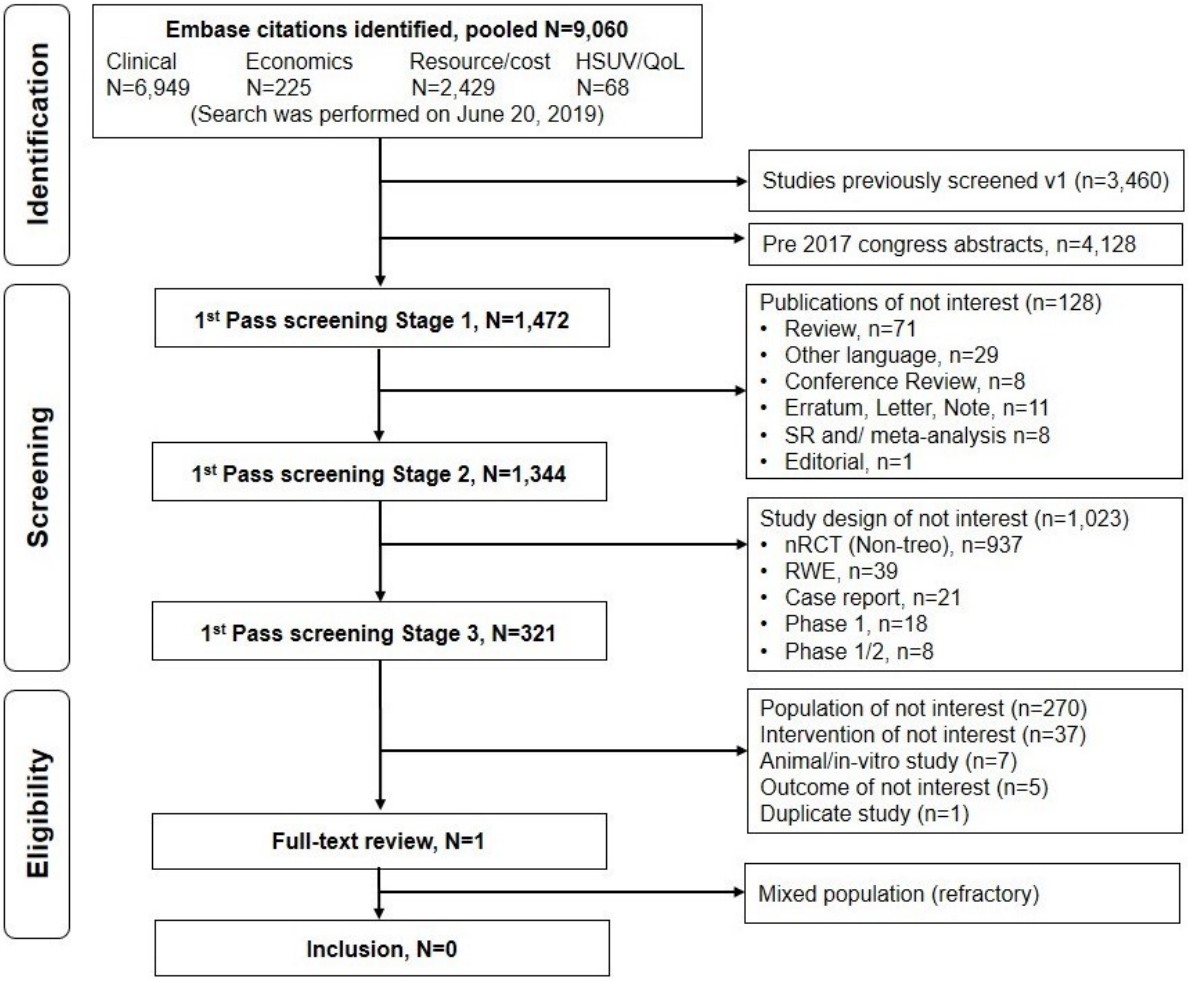
A2. A number of search terms are missing from the Embase searches which have been included in PubMed and Cochrane searches. The population facet for the

Embase searches only include terms for the most common malignant diseases, AML, ALL, MDS and MDS/MPN, while the Cochrane and PubMed searches include a far wider range of terms to cover malignant diseases. The HSCT facet in the Embase strategies for clinical effectiveness, cost use and utilities are missing terms like “marrow transplantation”, “umbilical cord blood transplantation” and “allogenic transplant” which have been included in PubMed and Cochrane searches. In addition, a range of intervention terms are missing from Embase searches. These include “total body irradiation” which is listed as a comparator in the final scope. The brand name for treosulfan, “ovastat” has not been searched for and neither have alternative spellings for treosulfan such as “threosulphan” and “tresulfan”. There are also terms used in PubMed and Cochrane searches to cover dose intensity which have not been included in Embase clinical effectiveness searches. Terms to cover standards of care and clinical management have been searched for in PubMed and Cochrane searches but have not been included in any Embase searches.

Please justify the use of these pared-down strategies in Embase which is the largest and most comprehensive of the databases searched. Please consider re-running the Embase searches with the same search terms which have been included in the PubMed and Cochrane searches to ensure that nothing has been missed.

We have performed additional targeted searches in Embase using the keywords / strings missing from the original EmBase searches to align them with the other database PubMed and Cochrane searches. The search strings for these additional searches are given in Table S1 of the Appendix. A total of 9,060 hits were identified and out of these 9,059 citations were excluded on Title/Abstract screening using the inclusion/exclusion criteria for the original systematic literature review (SLR). One full text paper (Maschan 2016) was then evaluated against the criteria and was excluded on the basis of ‘mixed population’ (36.4% of patients refractory to induction therapy). No additional studies were identified for inclusion into the systematic review. The following flowchart (Figure 1) shows the screening pathway for the additional searches.

Figure 1: Screening pathway for the additional searches



In a separate search we searched for 'ovastat' and found 53 hits. None of the hits was found suitable for inclusion other than those already included in the SLR. The alternative terms, e.g. treosulphan and tresulfan are included in the Emtree terms under 'Treosulfan'/exp used in the Embase searches. The term 'Threosulphan', when searched separately, generated 0 (zero) hits on Embase.

A3. #2 of the Embase searches has only a UK spelling for "leukaemia" and does not include the alternative "leukemia".

Please confirm if this is a reporting error and if not what effect this may have had on the overall recall of results.

The EmTree for 'leukemia'/exp includes leukaemia as one of the alternative spellings. This has also been verified by separately using 'Leukaemia' and 'leukemia'

in search strings and showing that the number of hits is identical. Thus omitting the alternative spelling for leukaemia would not affect the results of the SLR.

A4. #1 of the PubMed strings has used the UK spelling for "leukaemia" in their MeSH searches, e.g. "acute myeloid leukaemia [MeSH terms]". Unlike some interfaces, PubMed will not map incorrectly spelt MeSH terms to the correct MeSH term and as #1 has been searched in a block, it is possible that only the [Title/Abstract] searches have been included in the items found. In addition, Title/Abstract searches in PubMed only include spelling for "leukaemia" and not "leukemia". As terms for "leukaemia" have been entered twice for each disease, it is possible that this is a reporting error.

Please confirm if this is the case and, if not, please state what effect this may have had on the overall recall of results.

We checked the PubMed MESH Terms for 'Leukemia' as well as 'Leukaemia' – both point to the same MESH. Further 'Leukemia' as well as 'Leukaemia' when searched in 'All Fields' gives the same number of hits, i.e. 316,635 (25-6-19).

The duplicate entries in the searches look like a reporting error. To verify this, we performed the searches again using the corrected syntax (after removing 'a' from the duplicate string), we found similar number of hits i.e. 2,12,432 (24-06-19) to that reported in the SLR report, i.e. 2,12,253 (12-02-19). In view of this, there appears to be no possibility of missing any relevant references.

A5. Searches of Embase include a search for "refractory NEAR/2 (an$emia OR cytop$enia)". A search for this form of MDS has not been included in PubMed and Cochrane searches.

Please state what effect this may have had on the overall recall of results.

'NEAR/2' type of syntax is not available for PubMed searches. Thus this logic was not used while performing searches on PubMed. In the final inclusion/exclusion criteria used for the SLR, "refractory" was an exclusion criterion. Thus citations

identified by this term of the Embase search would have been excluded during screening. Omission from the PubMed searches would not have affected the overall results.

A6. Please confirm which databases were searched in the Cochrane Library.

The following databases were searched: Cochrane Reviews, Cochrane Protocols, Trials, Clinical answers, Editorials and Special collections, without using any filters.

A7. Please provide details of the source or reference for the filters used in clinical effectiveness, cost effectiveness and utilities searches.

Search terms to identify relevant publications for each element of the SLR were devised to meet the specific requirements of the SLR. No published filters were used.

Section B: Clarification on effectiveness data

The decision problem

B1. Priority question: Table 1 in document B of the CS specifies the population for this submission as "Adults, children and young people with malignant disease that is in remission before allogenic haematopoietic stem cell transplantation," in line with the final scope from NICE. However, the position of treosulfan in the treatment pathway is described (in section B.1.3 and elsewhere in the CS) as follows: "Treosulfan is a reduced-toxicity conditioning (RTC) regimen that aims to provide the efficacy of MAC regimens to patients who cannot tolerate such MAC regimens and would otherwise be treated with less efficacious RIC regimens i.e. patients aged ≥ 50 years at transplant and/or with a Haematopoietic Cell Transplantation-Comorbidity Index (HCT-CI) score >2 (indicating an intermediate to high risk of co-morbidities which influence NRM)." This latter definition is consistent with the

key evidence presented in the clinical effectiveness section of the CS and used to inform cost-effectiveness modelling.

Please clarify whether the intended population for this submission is as described in the scope or the more restricted population, as defined above and in-line with the evidence presented.

We have presented in our submission the best available evidence for treosulfan and this is based on the pivotal phase III trial (MC-FludT.14/L Trial II). This trial, with 570 randomised patients, was the largest ever prospective randomised clinical trial (RCT) comparing two conditioning regimens.[2] This trial focusses on "patients who cannot tolerate such MAC regimens and would otherwise be treated with less efficacious RIC regimens i.e. patients aged ≥ 50 years at transplant and/or with a Haematopoietic Cell Transplantation-Comorbidity Index (HCT-CI) score >2 (indicating an intermediate to high risk of co-morbidities which influence NRM)."

However, we believe the evidence from this population is applicable to the broader population of "Adults, children and young people with malignant disease that is in remission before allogenic haematopoietic stem cell transplantation" in line with the final scope. In Appendix L we provided the "Haematologists' Position Paper" from Prof. Suttorp et al. This position paper includes the opinion of six European KOLs (including 2 UK KOLs) and states:

"The signatories of this consensus statement wish to share their opinions and strong belief that TREO-based conditioning should not only be reserved to treat defined malignancies such as AML or MDS in certain age groups. We consider the existing trial data as compelling to support the use of TREO in most patients with malignant diseases requiring treatment with myeloablative conditioning followed by alloHSCT."

Therefore, we feel justified in believing that the trial population can be broadened in line with the Final Scope.

B2. Priority question: Table 1 in document B of the CS lists the relevant comparators as: cyclophosphamide and total body irradiation; cyclophosphamide and busulfan; busulfan with fludarabine; established

clinical management without treosulfan with fludarabine. This list is described as being "as final scope," however, a number of specific comparators listed in the scope appear to have been omitted: cyclophosphamide and thiotepa; high-dose busulfan with thiotepa; melphalan plus fludarabine. Furthermore, "established clinical management without treosulfan with fludarabine" is not listed in the scope.

a. Please provide a justification for these omissions as well as for the addition.

Because of the short timescale from referral to submission, without any prior warning, it was necessary to commission the work required for the submission on the basis of the post-referral Scope, and we were not aware that a revised Final Scope had been issued. Table 1 in document B should list the comparators as the final scope but with the removal of thiotepa (as discussed in our response to question A1).

Our submission focusses on reduced intensity conditioning because this is where we have direct phase III clinical trial evidence and also where there is an unmet need for a large proportion of patients. Further evidence for the comparison of treosulfan with fludarabine against cyclophosphamide, high-dose busulfan with thiotepa and melphalan plus fludarabine is based on the registry analysis which medac commissioned. The European Society for Blood and Marrow Transplantation (EBMT) registry was analysed and the results were discussed in section B.2.2.2.4 of the submission (a PDF is now provided as requested in question D2).[3]

In addition, the US Center for International Blood and Marrow Transplant Research (CIBMTR) was also recently asked by medac to analyse their registry and compare it to the data from the MC-FludT.14/L Trial II.[4] The CIBMTR report provides a comparison of allogeneic transplantation outcomes for acute myeloid leukemia (AML) and myelodysplastic syndrome (MDS) with full or reduced intensity busulfan with fludarabine regimens and busulfan-cyclophosphamide regimen in the United States to the results of the MC-FludT.14/L Trial II.

Lastly we refer you to the Haematologists' Position Paper within Appendix L in which six European Key Opinion Leaders summarise that "TREO/FLU conditioning leads to

better outcomes than busulfan-, cyclophosphamide, as well as melphalan-based regimens."

b. Did the systematic literature review identify any relevant studies for these comparators?

A number of studies with these combination therapies were captured in the searches, e.g. Cyclo-Thio (6 studies); BU-Thio (>10 studies); and Mel-Flu (>10 studies). Only RCTs based on intervention with the above combinations were eligible for inclusion. Most of these studies did not qualify for the final inclusion as per the eligibility criteria, except for one study on BU-Thio, i.e. Bacigalupo (2018) that was included in the SLR report.

c. Please confirm that the analysis of registry data (described in section B.2.2.4 of the CS) represents the only data being submitted on the comparative effectiveness of treosulfan with fludarabine vs. conditioning regimens other than busulfan with fludarabine.

As well as the EBMT registry data analysis comparing the effectiveness of treosulfan with fludarabine vs. conditioning regimens other than busulfan with fludarabine, we also have the CIBMT data which is described above. As mentioned above, we have also provided opinion from European KOLs (Appendix L: Haematologists' position paper) who have also reviewed the available data and confirmed their opinion that "TREO/FLU conditioning leads to better outcomes than busulfan-, cyclophosphamide, as well as melphalan-based regimens."

d. Please include a comparison between treosulfan with fludarabine and all comparators listed in the scope in terms of both clinical effectiveness and cost effectiveness.

Based on the systematic literature review (Appendix D), a feasibility analysis was performed to assess whether network meta-analyses could be performed to provide comparative evidence for treosulfan with fludarabine (TREO/FLU) versus other conditioning regimens of interest based on the NICE scope (Appendix L).

The feasibility assessment findings indicated that it is not possible to perform a network meta-analysis for any of the efficacy endpoints reported in the pivotal phase III trial.

Similarly, a feasibility assessment (Appendix L) for the completion of indirect and mixed treatment comparisons was completed based on the systematic literature review (Appendix D). The feasibility assessment considered whether indirect comparisons could be performed to provide comparative evidence for TREO/FLU versus other conditioning regimens of interest based on the NICE scope (Appendix L).

It was concluded that whilst indirect comparisons may be possible for TREO/FLU versus busulfan/cyclophosphamide (BU/Cy) and busulfan with fludarabine (BU/FLU; MAC) at 2 years for overall survival (OS), relapse rate (RR) and the incidence of graft-versus-host disease (GvHD), these outcomes are unlikely to provide sufficient, reliable and relevant comparative data for inclusion in the economic assessment of TREO/FLU as a conditioning regimen for patients undergoing HSCT as treatment for malignant disease.

Systematic review

B3. Appendix D of the CS: The reported objective and inclusion/exclusion criteria indicate that the systematic literature review was limited to patients with one of four categories of haematological malignancy (AML, ALL, MDS, MM).

Please confirm whether other haematological malignancies were excluded and, if yes, please provide the reason for exclusion.

According to the last EBMT report[5] in 2016, a total of 14,260 alloHSCTs were performed for malignant indications. From these, 11,423 were performed in patients with AML (n = 6281), MDS (n = 1878), ALL (n = 2651) and MM (n = 433). Therefore, these diseases are representative for alloHSCT indications (~79% of all alloHSCTs in Europe in 2016) and the literature search concentrated on these malignancies.

In addition, we believe that the effects of treosulfan are unrelated to the underlying condition as was noted in the Position Paper from Prof. Suttorp et al in Appendix L of our submission. This Position Paper is the “Consensus reasoned opinion by haematologists on the role of Treosulfan for patients with malignant diseases undergoing allogeneic stem cell transplantation”. Within this Position Paper, six European key opinion leaders (KOLs) (including 2 UK KOLs) and confirmed their opinion that the “advantages of TREO exist independently from the underlying disease to be treated.” In this position paper they explain how studies across adults with AML/MDS, AML, MDS, lymphoid malignancies, chronic myeloid leukaemia and multiple myeloma have demonstrated overall survival and NRM which was “mostly superior or at least as good as other RIC or MAC regimens” in their study of the literature.

B4. The methods section of appendix D of the CS indicates that 1st pass inclusion screening (titles and abstracts) and data extraction were completed independently by two reviewers.

Please confirm whether this was also the case for full text inclusion screening and assessment of the methodological quality of included studies. In addition, if quality assessment of relevant evidence was also performed using tools (e.g. Cochrane risk of bias tool, given that the only risk of bias assessment included in document B of the CS was for an RCT) other than the one based on the GATE tool (Appendix F of the process and methods guide (PMG4), please include this.

We confirm that 1st pass inclusion screening (titles and abstracts), data extraction and full-text review were completed independently by two reviewers.

A total of 21 clinical studies (9 RCTs, 12 observational — 6 retrospective and 6 prospective) was assessed for quality appraisal using National Institute for Health and Care Excellence (NICE) recommended questionnaires for quantitative intervention (https://www.nice.org.uk/process/pmg4/chapter/appendix-f-quality-appraisal-checklist-quantitative-intervention-studies). The quality assessment was initially completed by one analyst and reviewed by a senior reviewer.

B5. The results section of appendix D of the CS states that: “Eighteen studies of treosulfan-based RIC regimens were identified, including six publications reporting the results of clinical trials (phase II and phase III studies), four further prospective studies and eight retrospective studies.”

Please indicate which of the publications listed relate to the studies included in the main body of the CS (document B, section B.2, Clinical effectiveness). Please also explain what criteria were used to determine which of the studies identified in the systematic literature review should be included in document B.

Our submission document B included those trials sponsored by medac along with any publication from the studies. The trials included in document B are:

- The pivotal study MC-FludT.14/L Trial II which has been published (in abstract) form by Beelen et al in 2017[6] and 2019[7]. As these abstracts provide a sparse summary of the clinical data, we utilised the clinical study report (CSR) when putting together our submission.
- The MC-FludT.6/L study. This study was reported by Casper et al in 2010.[8] but was excluded from the systematic literature review (SLR; Appendix D) because the results provide for a mixed population and included those with acute myeloid leukaemia (AML), multiple myeloma (MM), myelodysplastic syndrome (MDS), chronic lymphatic leukaemia (CLL), chronic myeloid leukaemia (CML), Hodgkin’s disease (HD), acute lymphatic leukaemia, predominantly nodal lymphoma, primary extranodal lymphoma). Of these, patients with CLL, CML, HD and all types of lymphomas were an exclusion criteria.
- The MC-FludT.7/AML study. This study was reported by Casper et al 2012[9] and was identified in the SLR.
- The MC-FludT.8/MDS study. This study was reported by Ruutu et al in 2011[10]. This study included patients with refractory anaemia or refractory cytopenia subtypes of MDS and was therefore excluded from the SLR.

The SLR (Appendix D) identified five studies that reported the clinical outcome of treosulfan-based conditioning regimens as follows:

- Casper et al 2012[9]
- Kroger et al 2015[11]
- Deeg et al 2018[12]
- Kalwak et al 2018[13]
- Nemecek et al 2018[14]

These were all phase II studies and were not reported in the main submission except for Casper et al 2012 which was based on the medac study MC-FludT.7/AML. The Kalwak et al 2018 study [13] was an abstract publication only and therefore not reported in the main submission. In addition, Kroger et al 2015[11] was a multi-centre study without a comparator and investigated only 50 adult patients whilst Nemecek et al 2018[14] was also multi-centre and looked at just 40 children and young adult patients. The Deeg et al 2018[12] phase II study was randomized and included 100 patients (51 with MDS/chronic myelomonocytic leukaemia and 49 with MDS). We did not think that these smaller phase II studies added to the submission over and above the pivotal trial, MC-FludT.14/L Trial II,[2] which was the largest prospective randomised controlled trial comparing two conditioning regimens ever performed with 570 randomised patients.

Included trials: patient characteristics

B6. Priority question: As highlighted before (see question B1), key evidence presented in the clinical effectiveness section of the CS which is also used to inform cost-effectiveness modelling appears to be in a population more restricted than the population defined in the final scope.

Please confirm that you are not aware of any studies of the comparative effectiveness of treosulfan with fludarabine vs. other conditioning regimens for malignant disease that is in remission before allogenic haematopoietic stem cell transplantation that have been conducted in

a. **paediatric populations,**

b. **adults with haematological malignancies other than AML and MDS (those included in the key trial MC-FludT.14/L), and**

c. **adults who would be eligible for standard MAC (patients included in the key trial MC-FludT.14/L "were not eligible for a standard MAC busulfan- or TBI-based regimen").**

We confirm that no other prospective randomised studies with treosulfan-based conditioning versus other conditioning regimens are available. However, the following publications report retrospective comparisons of treosulfan-based conditioning versus other conditioning regimens in patients with various malignant diseases:

- Yerushalmi et al.(2015)[15] compared conditioning with treosulfan with fludarabine (TREO/FLU) to busulfan with fludarabine (BU/FLU) or fludarabine with fludarabine with melphalan (FLU/MEL) in patients with lymphoid malignancies. The study included 50 patients with various lymphoid malignancies given TREO/FLU and alloHSCT on a phase II prospective study. Outcomes following HSCT were compared retrospectively with an historical control group of 94 patients given a reduced-intensity conditioning (RIC) regimen of BU/FLU (FB2) or FLU/MEL before HSCT, over a 13-year period. There was no difference between the groups in median age, gender, donor, HCT-CI, disease type, number of prior therapies and status at stem cell transplant. Three-year OS was 67, 74 and 48% after TREO/FLU, FB2 and FLU/MEL in chemo-sensitive disease (P = 0.14) and 34%, 11% and 17% in chemo-refractory disease (P = 0.08). Three-year non-relapse mortality (NRM) was worst in the FLU/MEL group, whereas relapse mortality was comparable between the TREO/FLU and FLU/MEL regimens. Multivariate analysis identified a high comorbidity-score, chemo-refractory disease and FLU/MEL conditioning as associated with shortened survival. The authors concluded that FB2 is associated with low NRM and good results in chemo-sensitive disease, but with higher relapse mortality rates. FLU/MEL controls disease better, but with high NRM. TREO/FLU probably balances these outcomes more optimally. It is as safe as FB2 and as cytoreductive as FLU/MEL, resulting in improved outcome, especially in advanced disease.

- Shimoni et al. also in 2015[16] performed a retrospective analysis of all alloHSCTs for MDS performed between 2000 and 2011 and reported to the chronic malignancies working party of the EBMT (n = 2,516). They identified 480 patients given treosulfan with fludarabine (TREO/FLU) and compared their outcomes to patients given various MAC (n = 1,090) and RIC (n = 946)

regimens. TREO/FLU and RIC recipients were older than MAC recipients, median age 59, 60 and 50 years, respectively (P = 0.001). They were more likely to have an unrelated donor (P = 0.001) and peripheral blood stem cell grafts (P = 0.001). More TREO/FLU recipients had untreated MDS (33%, 20% and 24%) while less had chemosensitive disease (42%, 51% and 51%) (P = 0.001). More TREO/FLU recipients had prior transformation to AML (15%, 9% and 9%; P = 0.03), indicating an increased risk for the TREO/FLU group. Overall survival at 5 years after TREO/FLU conditioning in 480 MDS patients was 47% (95% CI 41-52%) and better than the respective values for RIC (39%; 95% CI 34-43%) and MAC regimens (38%; 95% CI 33-42%). Multivariate analysis identified age > 55 years (HR 1.7, P = 0.01), marrow blasts at HSCT > 10% (HR 1.5, P = 0.02), prior transformation to AML (HR 1.7, P = 0.01) and chemo-refractory disease (HR 1.8, P = 0.001) as poor prognostic factors for OS, while TREO/FLU-conditioning was protective (HR 0.6, P = 0.02). The authors concluded that TREO/FLU is associated with similar low relapse rates as MAC and similar low NRM as RIC, resulting in improved outcome over both RIC and MAC. TREO/FLU might be the preferred regimen for SCT in MDS.

- In a subsequent publication, Shimoni et al.[17] retrospectively compared HSCT outcomes after fludarabine with either intravenous busulfan at a myeloablative dose (FB4, 12.8 mg/kg, n = 1,265) or a reduced dose (FB2, 6.4 mg/kg, n = 1456) or treosulfan at 42 g/m² (TREO/FLU 14, n = 403) or 36 g/m² (TREO/FLU 12, n = 168). There was no difference in any outcome among patients in first complete remission at transplantation. However, there was better survival with treosulfan-based conditioning in advanced leukaemia. Furthermore, treosulfan-based regimens were associated with lower rates of graft-versus-host disease.

- Another EBMT group retrospectively compared the outcome of TREO/FLU, thiotepa-BU/FLU (TBF), and sequential fludarabine, intermediate dose Ara-C, amsacrine, total body irradiation/busulfan, cyclophosphamide (FLAMSA) conditioning in patients with refractory or relapsed AML. Overall survival at

2 years was highest with TREO/FLU (37%), and 24% for TBF, and 34% for FLAMSA.[18]

- Gran et al.[19] presented a large retrospective EBMT registry study in 4,544 multiple myeloma (MM) patients undergoing alloHSCT between 2008 and 2016. Treosulfan-based conditioning was received by 537 patients and OS, relapse-free survival (RFS), relapse incidence and NRM was compared to 2,830 non-treosulfan-RIC and 1,177 non-treosufan-MAC treated patients. Heterogeneity in the material makes it difficult to interpret results in the patients transplanted late in the course of the disease. Upfront alloHSCT was conducted in 1,103 patients and the 5-year OS in upfront Treosulfan-conditioned patients was 62%, which was significantly superior to both non-treosulfan-RIC and non-treosulfan-MAC patients respectively, apparently due to a tendency for lower NRM (10%) albeit a higher relapse rate. In multivariate analysis of treosulfan-based conditioning regimens, upfront line of conditioning was superior for OS, PFS and relapse incidence.

Paediatric studies

Furthermore, based on recent (unfortunately not yet fully published) study results, the current EWOG-MDS Guideline for HSCT in childhood MDS and juvenile myelomonocytic leukaemia (JMML) recommends the use of a conditioning regimen with treosulfan with fludarabine (TREO/FLU) ± thiotepa in most subtypes. Refractory cytopenia of childhood (RCC) is the most common subtype of paediatric MDS.[20]

HSCT following the previously recommended RIC preparative regimen with thiotepa plus fludarabine in 169 patients with hypocellular renal cell carcinoma (RCC) without karyotypic abnormalities resulted in a probability of overall and event free survival of 0.94 (95% CI, 0.90-0.98) and 0.83 (0.76-0.89), respectively. However, about 10% of patients experienced primary or secondary graft failure and 12% received a 2nd procedure such as a stem cell boost and/or 2nd HSCT.[20]

In contrast, all patients with RCC who have been transplanted following a treosulfan-based regimen (thiotepa/TREO/FLU) based on cellularity and/or karyotypic abnormalities experienced a prompt initial engraftment with a very low incidence of secondary graft failure (2/44 patients) and had a comparable overall outcome (OS

0.92 [0.83-1.00], EFS 0.86 [0.73-0.99]).[20] In light of these results it is consensus to transplant patients with hypocellular RCC without karyotypic abnormalities following a regimen consisting of treosulfan with fludarabine, aiming for an improved rate of engraftment.

Gonadal impairment is an important late effect with a significant impact on quality of life of transplanted patients. The aim of a recently published study was to compare gonadal function after busulfan or treosulfan-conditioning regimens in pre- and postpubertal children. This retrospective, multicenter study included children transplanted in paediatric EBMT centres between 1992 and 2012 who did not receive gonadotoxic chemoradiotherapy before the transplant.[21] 137 patients transplanted in 25 paediatric EBMT centres were evaluated. Median age at transplant was 11.04 years (range, 5 to 18); 89 patients were boys and 48 girls. Eighty-nine patients were prepubertal at transplant and 48 postpubertal. One hundred eighteen children received busulfan and 19 treosulfan. A higher proportion of girls treated with treosulfan in the prepubertal stage reached spontaneous puberty compared with those treated with busulfan ($P = 0.02$). Spontaneous menarche was more frequent after treosulfan than after busulfan ($P < 0.001$). Postpubertal boys and girls treated with treosulfan had significantly lower luteinizing hormone levels ($P = 0.03$ and $P = 0.04$, respectively) compared with the busulfan group. Frequency of gonadal damage associated with treosulfan was significantly lower than that observed after BU.[21]

B7. Priority question: Compared to patients seen in practice in England and Wales, how similar does the company consider the patients in the key trial MC-FludT.14/L (conducted in France, Germany, Hungary, Italy and Poland) to be? Have any clinical experts commented on this issue? If so, please provide relevant documents.

The general practice of alloHSCT is not different in England and Wales versus other major European countries (including France, Germany, Hungary, Italy and Poland). This is also demonstrated in the most recent EBMT report which shows that 52 UK teams performed 4,316 alloHSCTs in 2016.[5] Since these UK transplant sites are

members of the EBMT, they also treat their patients according to the EBMT Guidelines which are provided in the EBMT Handbook, version 2019.[22]

According to the latest British Society of Blood and Marrow Transplantation (BSBMT) registry (2017),[23] the most common indications for an alloHSCT in the UK are acute myeloid leukaemia (AML; 36%), acute lymphoblastic leukaemia (ALL; 16%) and myelodysplastic syndrome (MDS) or myelodysplastic/ myeloproliferative neoplasms (MDS/MPN; 13%). These indications for alloHSCT in the UK are comparable to those in adults and children undergoing HSCT in Europe (AML 38%; ALL 16%; and MDS or MDS/MPN 15%).[5] Further details were provided in section B.1.3.1. (Common haematological indications for HSCT in adults and children in the UK and section) B.1.3.1.3 (Patients requiring HSCT in the UK).

Within our submission in Appendix L we provided the "Haematologists' Position Paper" from Prof. Suttorp et al. As mentioned earlier, this Position Paper is the "Consensus reasoned opinion by haematologists on the role of Treosulfan for patients with malignant diseases undergoing allogeneic stem cell transplantation". Contributing experts for this document included also two transplant experts from the UK (Robert Wynn, Amit Patel). In their overall conclusion, these experts have stated that treosulfan-based conditioning should be made available for all patients with malignant diseases where an alloHSCT is indicated.

B8. Priority question: Different exclusion criteria and different treatment schedules (Tables 12 and 13 in the CS) were used in France to those used in the other study countries.

Please indicate whether any differences in the study results were observed (for any outcome measure) between geographic locations. Please provide the results for any subgroup analyses (by country) conducted.

In France, a different ATG preparation and regimen was used in matched unrelated donor (MUD) transplantations (ATG-Thymoglobuline i.v., 2.5 mg/kg on Day -2 and -1). This regimen is registered and was considered equivalent to the ATG preparation (ATG Fresenius) and the regimen used in the other countries.

In France, no AML low risk and MDS very low/low risk were included due to a request from the French authorities. In the total study population, only 10.5% of low risk AML were included and 20.1% of very low/low risk MDS.

Furthermore, only 14 patients were recruited from France into this study (~ 2.5% of all randomised patients). This low number could not compromise the overall study results and a subgroup analysis for such a small patient number would not be useful.

A country-specific sub-analysis of treatment results was not performed.

B9. Please confirm that the analysis of registry data (described in section B.2.2.4 of the CS) represents the only data being submitted on the comparative effectiveness of treosulfan with fludarabine vs. conditioning regimens other than busulfan/fludarabine.

Please see response to B2c.

Included trials: efficacy results

B10. Section B.2.13 of document B of the CS includes the following statement: "In addition, a clinically relevant positive effect of treosulfan on patient functional ability, status of general well-being and activities of daily life was shown with treosulfan. Post-surveillance data demonstrated the clinically relevant long-term benefit of treosulfan up to 4 years after HSCT."

Please provide numerical results in support of this statement, e.g. the last post-surveillance time point reported in section B.2.6 is 36 months.

The primary endpoint of the trial was EFS at 24 months after transplant and our data for Karnofsky Performance Status (KPS), a measure of the functional status of patients, was also only measure for 24 months. However, in the post-surveillance evaluation, follow-up data for EFS, OS, relapse/progression, and NRM were obtained for a period of up to 4 years after transplantation. The median (range) follow-up of patients was 29.7 (3.0, 52.1) months in the treosulfan group and 29.4 (3.9, 54.3) months in the busulfan group.

The Kaplan-Meier estimates of EFS, OS, and NRM in the clinical study report show the survival curves up to 51 months after transplant and demonstrate that the advantage of treosulfan is maintained over this time period (i.e. > 4 years = 48 months).[2] Submission B provides the Kaplan-Meier graphs of EFS (Figure 4). Below are the Kaplan-Meier estimate of OS (Figure 2), and the cumulative incidence of relapse/progression (Figure 3).

Figure 2: Kaplan Meier estimates of OS (Full Analysis Set) – MC-FludT.14/L Trial II

Source: MC-FludT.14/L Trial II[2]
Abbreviations: CI, confidence interval; FAS, full analysis set
a Adjusted for donor type as factor, and risk group and centre as strata using Cox regression model.
b For testing non-inferiority of treosulfan compared to busulfan.
c For testing superiority of treosulfan compared to busulfan.

Figure 3: Cumulative incidence of relapse/progression (Full Analysis Set) – MC-FludT.14/L Trial II

Source: MC-FludT.14/L Trial II[2]
Abbreviations: CI, confidence interval; NA = Not applicable.
[a] Adjusted for donor type as factor and risk group as stratum using Fine and Gray model
[b] Based on test of Gray.

Ongoing studies

B11. Section B.2.11 of document B of the CS lists one ongoing study (MC-FludT.16/NM); the population of this study is outside the final scope for this appraisal (paediatric patients with non-malignant disease). A further study (MC-FludT.17/M) is described elsewhere (section B.2) as "long-term follow-up ongoing".

Please confirm that you are not aware of any further ongoing studies of treosulfan or any of the comparators listed in the final scope.

There are no other ongoing treosulfan studies sponsored by medac other than MC-FludT.16/NM and MC-FludT.17/M.

However, we are aware of a number of ongoing investigator-initiated trials with treosulfan which are summarised below.

EudraCT Number:	2012-005414-18
Sponsor:	University Medical Center Freiburg / Germany
Full Title:	Thiotepa-Fludarabine-Treosulfan (TFT) conditioning for 2nd allogeneic PBSCT from a different unrelated donor in patients with AML relapsing from prior allogeneic HCT
Medical condition:	Patients with relapse of AML after the 1st allogeneic PBSCT
Conditioning:	Thiotepa-Fludarabine-Treosulfan (TFT)
Population age	Adults
Link	https://www.clinicaltrialsregister.eu/ctr-search/search?query=eudract_number:2012-005414-18

EudraCT Number:	2012-003032-22
Sponsor:	St. Anna Kinderkrebsforschung (ALL-SCT-ped-FORUM study)
Full Title:	Allogenic stem cell transplantation in children and adolescents with acute lymphoblastic leukaemia - FORUM (For Omitting Radiotherapy Under Majority age)
Medical condition:	ALL
Conditioning:	TREO/FLU/ thiotepa or BU/FLU/ thiotepa versus TBI/etoposide
Population age:	Infants and toddlers, Children, Adolescents, Under 18, Adults
Link:	https://www.clinicaltrialsregister.eu/ctr-search/search?query=eudract_number:2012-003032-22

EudraCT Number:	2007-005477-54
Sponsor:	OSPEDALE S. RAFFAELE DI MILANO
Full Title:	Treosulfan-based conditioning and Rapamycin-based GvHD prophylaxis prior to un-manipulated allogeneic haematopoietic stem cell transplantation from a mismatched donor in patients with high risk haematological malignancies
Medical condition:	Neoplastic and haematologic pathologies
Conditioning:	TREO/FLU + total marrow irradiation
Population age:	Newborns, Infants and toddlers, Children, Adolescents, Under 18, Adults, Elderly
Link:	https://www.clinicaltrialsregister.eu/ctr-search/search?query=eudract_number:2007-005477-54

EudraCT Number:	Not available
ClinicalTrials.gov Identifier:	NCT00796068
Sponsor:	Fred Hutchinson Cancer Research Center
Full Title:	Treosulfan, Fludarabine Phosphate, and Total-Body Irradiation in Treating Patients With Hematological Cancer Who Are Undergoing Umbilical Cord Blood Transplant

Medical condition:	ALL, AML, CML, MDS
Conditioning:	Treosulfan + fludarabine + TBI
Population age:	up to 65 Years (Child, Adult, Older Adult)
Link:	https://clinicaltrials.gov/ct2/show/NCT00796068?term=treosulfan&recrs=ab&rank=6

Ongoing studies with comparators are listed below.

EudraCT Number:	2009-015968-34
Sponsor:	FRED HUTCHINSON CANCER RESEARCH CENTER / USA
Full Title:	Nonmyeloablative Conditioning with Pre- and Post-Transplant Rituximab followed by Related or Unrelated Donor Hematopoietic Cell Transplantation for Patients with Advanced Chronic Lymphocytic Leukemia
Medical condition:	Chronic lymphocytic leukemia (CLL)
Conditioning:	Nonmyeloablative conditioning, not specified
Population age:	Adolescents, Under 18, Adults, Elderly
Link:	https://www.clinicaltrialsregister.eu/ctr-search/search?query=eudract_number:2009-015968-34

EudraCT Number:	2009-014980-38
Sponsor:	CHU de Liège / Belgium
Full Title:	Co-transplantation of mesenchymal stem cells and HLA-mismatched allogeneic hematopoietic cells after reduced-intensity conditioning: a phase II randomized double-blind study.
Medical condition:	Haematological malignancies confirmed histologically and not rapidly progressing (AML, CML)
Conditioning:	RIC, not specified
Population age:	Children, Adolescents, Under 18, Adults, Elderly
Link:	https://www.clinicaltrialsregister.eu/ctr-search/search?query=eudract_number:2009-014980-38

EudraCT Number:	2017-004800-23
Sponsor:	University of Birmingham
Full Title:	A comparison of reduced dose total body irradiation (TBI) and cyclophosphamide with fludarabine and melphalan reduced intensity conditioning in adults with acute lymphoblastic leukaemia (ALL) in complete remisssion.
Medical condition:	Acute Lymphoblastic Leukaemia (ALL)
Conditioning:	TBI/CY versus FLU/MEL
Population age:	Adults
Link:	https://www.clinicaltrialsregister.eu/ctr-search/search?query=eudract_number:2017-004800-23

EudraCT Number:	2012-005378-73
Sponsor:	FONDAZIONE NEOPLASIE SANGUE ONLUS
Full Title:	Allogeneic Transplantation after a Conditioning with Thiotepa, Busulfan and Fludarabin for the treatment of refractory/early relapsed aggressive B-cell non Hodgkin lymphomas: a Phase II Multi-Center trial.
Medical condition:	Refractory/early relapsed aggressive B-cell non Hodgkin lymphomas

Conditioning:	BU/FLU/ thiotepa
Population age:	Adults
Link:	https://www.clinicaltrialsregister.eu/ctr-search/search?query=eudract_number:2012-005378-73

EudraCT Number:	2018-000356-18
Sponsor:	University Medical Center Utrecht
Full Title:	Individualized dosing of fludarabine during innate allo SCT: A randomized phase II study (TARGET Study)
Medical condition:	AML, MDS, ALL, CML, CLL, NHL, HL, or a myeloproliferative disease (MPD)
Conditioning:	BU/FLU
Population age:	Adults, Elderly
Link:	https://www.clinicaltrialsregister.eu/ctr-search/search?query=eudract_number:2018-000356-18

EudraCT Number:	2012-005538-12
Sponsor:	University of Birmingham
Full Title:	Randomised Trial of the FLAMSA-BU Conditioning Regimen in Patients with Acute Myeloid Leukaemia and Myelodysplasia Undergoing Allogeneic Stem Cell Transplantation
Medical condition:	AML, MDS
Conditioning:	FLAMSA-BU
Population age:	Adolescents, Under 18
Link:	https://www.clinicaltrialsregister.eu/ctr-search/search?query=eudract_number:2012-005538-12

Indirect comparisons/NMA

B12. Priority question: Section B.2.9 of document B of the CS reports that some indirect comparisons would be possible at 2 years for OS, RR and GvHD incidence.

Please perform these indirect comparisons and provide the data and results. If possible, please use the hazard ratio for OS if it is reported by all relevant studies.

As discussed in the response to B.2.d, an indirect comparison although technically possible for some of the endpoints was not thought to be sufficiently informative for the submission. The economic model is based on the pivotal trial (MC-FludT.14/L Trial II) and utilised the clinical endpoints of both OS and EFS. As the indirect comparison cannot provide data on EFS it has not been performed.

Section C: Clarification on cost-effectiveness data

Cost-effectiveness analyses

C1. Priority question: Please perform cost effectiveness analyses that include all comparators as specified in the scope. These analyses should use any relevant clinical effectiveness data (see question B2) and as requested in question B12.

Please refer to the responses to B2 and B12.

Clinical inputs

C2. Priority question: Please provide all details of the communication between the company and the clinical experts. Please include anonymised information about the clinical experts, detailed minutes of the face-to-face meeting and/or teleconference, list of expert recommendations and justifications for clinical assumptions and inputs used in the model. In particular, please indicate the following:

How many experts provided information for each of the following: model structure, modelling of OS, modelling of EFS, choice of the cure point, health state resource use and costs and health state utilities? In each case, please provide more detail of the clinical/working setting and experience of included experts.

Three clinical experts were consulted as part of the economic model development process; two UK based experts who are both employed by established Universities in England, with one expert holding a position as a Director within a hospital. The third expert consulted is an employee of Medac. This answer focus on the first two experts previously referenced and both of these experts hold the title of Professor who have considerable experience of the UK setting.

The two clinical experts were interviewed via teleconference and provided information relating to modelling of OS and EFS, particularly in regard to assumptions around long-term mortality and the use of a 5-year cure point. The first

clinical expert was interviewed for 2 hours, with the second clinical expert interviewed for 1 hour.

Both clinical experts confirmed that the hazard ratio of 2.3 from TA552 (sourced from Martin 2010) was the most plausible estimate from Table 1 of the interview guide for calculating long-term mortality of patients post-HSCT. Both clinical experts also confirmed that the 5-year cure point curves were the most suitable of the options presented, with excess mortality for post-HSCT patients compared to the general population.

The first clinical expert interviewed raised an initial concern around the slight overestimation of patients in the relapse health state in the long-term but agreed that the curves in Figure 6 of the interview guide (used in the base-case analysis) were still plausible following further discussion regarding the absolute differences between the curves and the approximate proportion of patients in the relapse state after 100 model cycles.

The second expert noted that other studies had been published on long-term mortality among patients undergoing HSCT with leukaemia, and specifically mentioned studies by Socie et al, Goldman et al and Gunnarsson et al. A study by Socie et al published in 1999 was identified[24] – this study looked at mortality for patients in continuous complete remission two years after transplantation. This study was considered to be less appropriate than the Martin 2010 study[25] as it focused on patients two years after transplantation (rather than the cure point of five years confirmed with clinical experts), and did not include patients with MDS. Following the interview, a 2010 study by Goldman et al[26] was identified that studied relapse and late mortality among patients surviving 5 years after receiving a myeloablative alloHSCT. However, this study focused exclusively on patients with chronic myeloid leukaemia, and thus was assumed to be less generalisable to the trial population than the Martin 2010 study (which included patients with AML and MDS). A study by Gunnarsson et al specifically focusing on mortality among patients undergoing HSCT could not be identified through targeted review of the literature, although a study of long-term survival among patients with chronic myeloid leukaemia by Gunnarsson et al was collected (but did not include estimates of long-term mortality specific to HSCT patients).[27]

Each clinical expert also provided commentary on the utility functions and values presented in the interview discussion guide. Both clinical experts interviewed agreed that the value set (value set 1) produced for post-HSCT recovery patients from the Grulke study using the Proskorovksy mapping algorithm was the most plausible of those presented, due to the more gradual improvement in quality of life following an HSCT and closer approximation to the general population utility for long-term survivors (year 4+ post-HSCT). Following discussion with the second clinical expert interviewed, it was agreed that application of the Year 4+ post-HSCT utility to functionally cured patients was more appropriate than general population utility, given the excess mortality assumed for patients in long-term recovery compared to the general population.

In terms of relapse utilities, both clinical experts agreed that the estimates produced for the relapse/progression utility for value set 1 were plausible, although the second expert interviewed noted that the quality of life of patients relapsing/progressing following an HSCT may be lower than the estimate from value set 1 given that patients relapsing following an HSCT generally have poor survival outcomes. As such, a scenario analysis was conducted where the relapse disutility multiplier was applied to the discharge health state utility rather than the ≤6 months post-HSCT recovery utility (Scenario 4, Section B3.6.3.4 of the original submission), which generated a lower utility estimate for relapse patients than for the base case analysis.

The first clinical expert interviewed provided information on the regimens that patients relapsing following an HSCT would receive, indicating that patients would receive either palliative care, salvage chemotherapy or hypomethylating agents if relapsing within one year, and either chemotherapy (FLAG + IDA) or a second HSCT. In both cases, the clinical expert was unsure what percentage of patients would receive each treatment but agreed that assuming an equal distribution across each treatment option was appropriate. For secondary HSCTs, the first clinical expert was unsure about what conditioning regimens patients would necessarily receive but agreed that assuming half of patients receiving treosulfan-based conditioning and half of patients receiving busulfan-based conditioning was an appropriate assumption. The first clinical expert also provided information on what

resource use patients in long-term recovery would need, indicating that patients in long term recovery (after 1 year) would still have follow-up visits and blood tests every 4-6 months. As prior cost estimates had been collected from TA545 for the first two years of post-HSCT recovery which were anticipated to provide more conservative results for treosulfan, this assumption was adopted for patients still remaining event free after two years, with costs for blood tests, biochemistry profile, phlebotomies and consultant-led clinical haematology appointments assumed to occur every four months until patients relapse/progressed or died.

The second clinical expert interviewed agreed that the assumptions around relapse treatment regimens and resource use among long-term post-HSCT recovery patients were appropriate.

b. **Please provide further details of the opinions given by experts in relation to each aspect of the model listed in part a. of this question and provide details regarding the extent to which these opinions were included in the model or justification of why they were not included.**

Please see the response to part (a).

c. **In particular, please provide the answers given by experts to the questions presented in Appendix L3 of the CS.**

Please see the interview guides attached as additional appendices with notes summarising the responses provided by each expert.

Model structure

C3. On page 52 in document B of the CS, event-free survival (EFS) is defined as follows: “Event-free survival within 2 years after transplantation was defined as the primary endpoint of the trial. EFS was measured from time of end of HSCT (= Day 0) to time of event. Events were defined as relapse of disease, graft failure or death (whatever occurred first)”.

Please explain how graft failure was exactly modelled and justify why it was not modelled as a health state.

Graft failures occurred in very few patients in the pivotal phase III clinical trial (MC-FludT.14/L Trial II). In both treatment arms, 1 patient (0.4%) experienced primary graft failure. 8 patients (2.8%) in the busulfan arm experienced secondary graft failure, with no patients experiencing secondary graft failure in the treosulfan treatment arm. As such, an extrapolation of long-term survival for graft failure patients would be unreliable given the very small sample size of this patient group (particularly for treosulfan).

Additionally, no data were identified in the systematic and targeted review of the literature that indicated differences between relapse/progression patients and graft failure patients in terms of either costs or quality of life.

Graft failure is captured as an “event” in terms of EFS, so graft failure patients are considered in terms of survival but were not differentiated from relapse/progression patients in terms of costs and non-survival outcomes in the economic model.

Treatment effectiveness – Long-term mortality

C4. Priority question: Please provide details on the feedback from the clinical expert(s) regarding the choice of the different SMRs for long-term mortality. This could be answered, for example, by adding more details to Table 41 in document B of the CS (e.g. explain the assumptions for each approach, is the population in TA552 comparable to the one in this appraisal – so that the same ratio can be used etc.).

Please see response to question C2.

C5. Priority question: Please answer the following questions regarding the third approach used to model long-term mortality (Parametric curves or HSCT-specific life tables (as determined by SMRs), depending on which has the highest mortality rate).

a. **The description of this method is unclear. This method seems to suggest that, unlike method 5, there is no fixed cure point, i.e. the cure point is determined by the time point where the HSCT-specific mortality rate is highest. Please clarify whether this is correct or not.**

The ERG's interpretation of the third survival analysis approach is correct. Selection of this approach will override the cure point assumption of 5 years.

b. **If a. is correct, please provide an approximate estimate of the cure point using method 3 and discuss its clinical validity.**

The cure points for each model engine are shown below in Table 1. This approach was applied to both OS and EFS curves. There were different cure points for each survival curve (especially for the Busulfan OS AML arm, where the SMR-adjusted life table mortality immediately exceeded that of the parametric survival curves). For this reason, this approach was not selected as the base case analysis. Its clinical validity was discussed with experts, who felt that selecting a defined cure point of 5 years would be the most appropriate course of action.

Table 1: Cure points when the third approach to survival modelling (choosing the highest rate of mortality using either the parametric models or HSCT-adjusted life tables)

Survival curve	Model arm	Approximate cure point, years
OS	Treosulfan: pooled cohort	6.8
	Treosulfan: AML only	6.6
	Treosulfan: MDS only	6.8
	Busulfan: pooled cohort	5.2
	Busulfan: AML only	4.4†
	Busulfan: MDS only	5.5
EFS	Treosulfan: pooled cohort	4.6
	Treosulfan: AML only	4.2
	Treosulfan: MDS only	5.8
	Busulfan: pooled cohort	5.1
	Busulfan: AML only	5.2
	Busulfan: MDS only	4.8

Abbreviations: AML, acute myeloid leukaemia; EFS, event-free survival; MDS, myelodysplastic syndrome; OS, overall survival.
†This estimate excludes the first model cycle, where the HSCT-adjusted life table mortality estimate is instantaneously higher than the mortality estimate from the parametric curve for the first 28-day period.

c. **If a. is correct, please explain the role of the parameter 'cure time' in the model. The model forces the user to choose a cure time, for example 5 years. If this is chosen as cure time, and method 3 results in a different (shorter) cure time, which one is valid and how is it used in the model?**

If method 3 is chosen, the cure time parameter has no effect in terms of the modelling of survival. In this circumstance, the dynamic cure time as defined by the formulae in columns S and AE in each of the engine worksheets would be used.

d. **Please consider adding the following method to the model, provided that it is different to method 3: Parametric curves or HSCT-specific life tables (as determined by SMRs), depending on which has the highest mortality rate, up to a 5-year cure point, followed by switch to HSCT-specific life table (as determined by SMRs) mortality rates. Note that this method considers that the cure point is allowed to happen before 5 years. When this happens, please note that the parameter cure time (the user can choose between 1 and 5 now) has to be consistent with the one provided by the curves. This is because the cure point is used for the calculation of utilities.**

The model was amended to utilise option 5, with the possibility to be considered functionally cured earlier than the 5-year cure point if the mortality rate on the EFS parametric curve exceeded that of the SMR-adjusted HSCT mortality. When patients were considered functionally cured, they would use the associated utility value for the general population.

The results generated for this method are identical as for the amended base case (see Question C21d, C24 and of the document for more information); none of the mortality rates generated by the EFS curves are lower than those generated using the HSCT-adjusted life tables within the 5-year parametric modelling period. Patients would need to be cured earlier than 4 years in order to generate alternative results.

C6. Page 126 in document B of the CS states "As HSCT is a potentially curative treatment, the option to select a 'cure point' in the model was implemented".

Furthermore, page 118 states “treosulfan is a conditioning therapy that impacts the effectiveness of HSCT”.

Please clarify whether this impact on the effectiveness of HSCT also affects treatment, i.e. please justify why the cure point is equal for both treatment arms.

Patients who survive alloHSCT for at least 5 years can be considered as cured. Relapses/transplant-related deaths after 5 years are very rare. This is also shown in the respective survival curves of our pivotal trial showing a plateau in both groups starting after approx. 39 months (Figure 2 provided earlier). Treosulfan’s impact on effectiveness is the about 10% higher cure rate.

C7. Page 126 in document B of the CS states “Prior to the cure point, the parametric curves for OS and EFS are used. After the cure point, mortality is determined by using life tables for the general population, or life tables with standardised mortality ratios (SMRs) for HSCT applied”.

Please clarify whether this approach implicitly assumed that the cure rate for HSCT is 100% (i.e. after the cure point – 5 years – all patients are cured). Please justify this assumption.

It should be noted that the term “cure point” refers to patients being functionally cured. There is still the possibility for patients to relapse to AML / MDS after the 5-year cut-off point, as the switch from parametric curves to SMR-adjusted life tables is applied to both OS and EFS curves. This was done to allow a small proportion of patients to relapse after the 5-year “cure point”, in line with KOL feedback.

The model approach used assumes that when patients have survived for 5 years post-HSCT, the majority of mortality risk for this population will not be due to potential relapse of AML and MDS, and will instead be due to other causes, for example, long-term complications arising from HSCT itself. Therefore it is no longer appropriate to use the parametric survival models for the portion of the time where HSCT patients will be at high risk of relapse and progression, i.e. in the years immediately following HSCT.

C8. On page 128 in document B of the CS, five different approaches to modelling long-term mortality are presented. Based on the information on pages 126 and 127, it seems that methods 1, 2 and 4 are implausible/ incorrect.

Please clarify whether this is the case.

Methods 1, 2 and 4 were included as potential options as they are required to derive the plausible scenarios (3 and 5), and were included in the model for transparency and to show how the Company arrived at the modelling approach which was used in the base case. Their inclusion in the model was also useful for internal validation.

Treatment effectiveness – Survival

C9. Priority question: Please provide the following information regarding both, EFS and OS.

a. **Log-log plots and assessment of proportional hazards.**

Please see log-log plots below for OS and EFS for the overall patient population (AML + MDS).

Figure 4: AML + MDS - OS

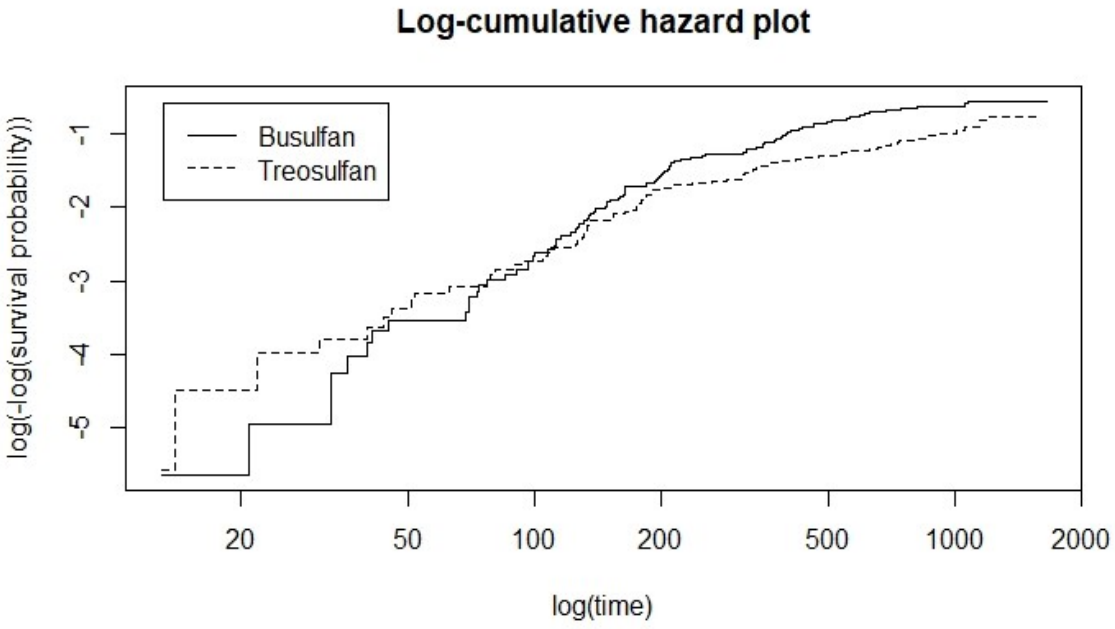
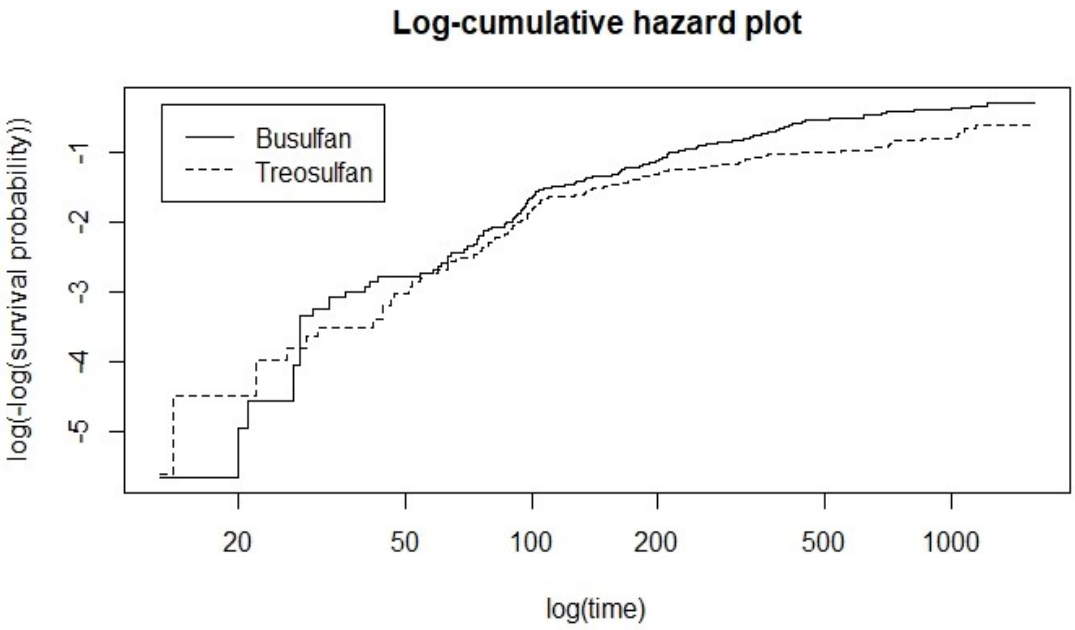


Figure 5: AML + MDS - EFS



For both OS and EFS, the curves for treosulfan and busulfan are broadly non-parallel and overlap, suggesting that assuming proportional hazards is not appropriate.

b. **To further assess proportional hazards, please consider (for example) the following methods: statistical test for non-proportional hazards (e.g. Therneau-Grambsch – standard in R), plot of trial-based hazard functions (per treatment arm) vs. time, plot of trial-based hazard ratio vs. time.**

Please see response to part (a). We believe that the log-log plots provide justification that assuming proportional hazards is not appropriate.

c. **If after the assessments in b. the proportional hazards assumption seems plausible (or unclear) please include proportional hazard extrapolations in the model.**

Please see response to part (a).

d. **For each parametric curve included in the model, please provide a plot vs. KM data.**

These plots are available in the economic model on the following worksheets:

- Treosulfan_EFS_inputs
- Busulfan_EFS_inputs
- Treosulfan_OS_inputs
- Busulfan_EFS_inputs

e. **For each parametric curve included in the model, please provide a plot of the hazard rate functions (per treatment arm) vs. time.**

The hazard rate functions can be derived from the survival functions presented in the EFS and OS input worksheets of the model shown in C9d. It is unclear how these plots would be informative for the purposes of this analysis.

f. **For each parametric curve included in the model, please provide a plot of the hazard ratio vs. time.**

As for question C9e, it is unclear how these additional plots would be of use for the analysis.

g. **Please provide probability estimates to validate the tails of the parametric curves presented in the electronic model (Treosulfan_EFS_inputs and Busulfan_EFS_inputs sheets). These estimates could be based on clinical evidence, the literature or expert opinion and should be given for the pooled, AML and MDS populations. Note with the current information, it is extremely difficult to select a curve based on solid arguments.**

The tails of the parametric curves were only validated up to the 5-year cure point, as the experts consulted felt that using the whole parametric curves to model survival was inappropriate. Both clinical experts interviewed agreed that the curve presented in Figure 6 produced a realistic transition in terms of switching from the parametric survival curve to the adjusted life table data.

h. **Please explain why "Flexible spline models were not included in the economic analysis due to concerns with over-fitting to the observed data". If this was based on specific analyses, please include them in the answer.**

No formal statistical analyses were performed to determine whether the flexible spline models over-fit. The parametric survival models and mixture-cure models appeared to fit the data sufficiently well, so the flexible spline approach was not pursued further given that it also did not appear to produce significant improvements in fit according to AIC statistics and visual inspection. The Company also feels that MCM / NMCM / parametric survival models are easier to understand conceptually. Flexible splines were not used in the base case analysis of any previous AML STAs and were only considered as scenario analyses. The cure function model approach appeared to be more appropriate for the base case analysis as HSCT is a potentially curative treatment, and given their use in a prior AML technology appraisal (TA545).

i. **Please explain how visual inspection and expert opinion were used to select parametric curves. The text in the CS is unclear and it seems the selection was solely based on lowest AIC.**

AIC was ultimately the deciding factor in selecting the parametric curves used. While MCM Weibull and NMCM Weibull models also produced reasonable fits based on visual inspection, NMCM lognormal and MCM lognormal consistently produced the lowest AIC estimates across EFS and OS curves for treosulfan and busulfan for the overall population. Limited differentiation between NMCM lognormal and MCM lognormal was possible on the basis of AIC and visual inspection, and as such NMCM lognormal estimates were selected on the basis of AIC.

j. **Please provide the underlying assumptions and the equations of the mixture-cure models (MCM) / non-mixture-cure models (NMCM) used in the CS.**

The mixture cure and non-mixture cure models follow the same principle where a proportion of the population are assumed to achieve a cure status (the cure fraction), with parametric survival models applied to the uncured fraction. The functional forms of the MCMs and NMCMs have the following functional forms:

MCM: θ + (1 - θ)*S(t)

NMCM: θ^(1 - S(t))

Where θ is the cure fraction and S(t) represents the chosen parametric survival function.

k. Please provide the code used to estimate all the parametric equations used in the CS.

R code used to estimate the parametric equations used in the original submission are attached as an additional appendix.

C10. Please provide the following plots.

a. OS KM curves vs. HSCT-specific survival data (as determined by SMRs).

b. For each parametric OS curve included in the model, please provide a plot of the survival function (per treatment arm) vs. HSCT-specific survival data (as determined by SMRs). Please provide these figures for the pooled AML+MDS, the AML-only and the MDS-only populations separately.

These plots are available in the economic model on the input sheets described in question C9d.

Subgroup analysis

C11. Please explain the rationale for including cost effectiveness subgroup analyses (AML/MDS) in the submission even though these were not mentioned in the scope. Furthermore, please clarify whether the subgroups considered (AML/MDS) are appropriate or not. In Section A.1 in Document A of the CS it is mentioned that the underlying disease is not the "primary determinant of conditioning regimen". Risk factors seem to be more important.

Individual subgroup analyses were performed as a validation exercise to ensure that results were broadly consistent for treosulfan within each patient indication, and that

results pooled from the individual indications were consistent with the overall results for the total AML and MDS population.

Costs and resource use

C12. Priority question: In Appendix I, there appears to be a discrepancy between Table 7 (which mentions a second HSCT only in late relapse / after one year) and Table 8 (which mentions a HSCT also in the first year / early relapse).

Please provide an explanation to resolve this (apparent) discrepancy, and clearly indicate whether or not a second, or even third, HSCT is possible only in late relapse or in early relapse as well.

HSCT costs were incorrectly included in the sections of Table 8 of Appendix I describing costs for patients relapsing in the first year post-HSCT. A corrected version is presented below with the row for HSCT costs removed for patients relapsing in the first year.

Table 7 (Appendix I): List of health states and associated costs in the economic model

Health states	Items	Value	Reference in submission
EFS: 6 to 12 months post-HSCT (cost per cycle)	Total cost per cycle	£4,959.47	**Error! Reference source not found.**
EFS: 7 to 12 months post-HSCT (cost per cycle)	Total cost per cycle	£3,407.04	**Error! Reference source not found.**
EFS: 12 to 24 months post-HSCT (cost per cycle)	Total cost per cycle	£1,228.97	**Error! Reference source not found.**
EFS: 24 months onwards post-HSCT (cost per year)	Total cost per cycle	£40.40	**Error! Reference source not found.**
Relapsed / progressed AML: Early relapse (≤12 months)	HCRU costs per cycle	£1,548.72	**Error! Reference source not found.**
	Monitoring costs per cycle	£38.75	**Error! Reference source not found.**
	Blood transfusion costs per cycle	£1,764.24	**Error! Reference source not found.**
	Relapse treatment (pharmacy) costs per cycle	£285.37	**Error! Reference source not found.**
	Total	£3,637.07	**Error! Reference source not found.**
Relapsed / progressed MDS: Early relapse (≤12 months)	HCRU costs per cycle	£1,732.04	**Error! Reference source not found.**
	Monitoring costs per cycle	£38.75	**Error! Reference source not found.**
	Blood transfusion costs per cycle	£1,764.24	**Error! Reference source not found.**
	Relapse treatment (pharmacy) costs per cycle	£285.37	**Error! Reference source not found.**
	Total	£3,820.39	**Error! Reference source not found.**
Relapsed / progressed AML: Late relapse (>12 months)	HCRU costs per cycle	£1,548.72	**Error! Reference source not found.**
	Monitoring costs per cycle	£38.75	**Error! Reference source not found.**
	Blood transfusion costs per cycle	£1,764.24	**Error! Reference source not found.**
	Relapse treatment (pharmacy) costs per cycle	£285.50	**Error! Reference source not found.**
	HSCT cost (one-off)	£46,413.11	**Error! Reference source not found.**
	Total (excluding HSCT)	£3,637.21	**Error! Reference source not found.**
Relapsed / progressed MDS:	HCRU costs per cycle	£1,732.04	**Error! Reference**

Health states	Items	Value	Reference in submission
Late relapse (>12 months)			**source not found.**
	Monitoring costs per cycle	£38.75	**Error! Reference source not found.**
	Blood transfusion costs per cycle	£1,764.24	**Error! Reference source not found.**
	Relapse treatment (pharmacy) costs per cycle	£285.50	**Error! Reference source not found.**
	HSCT cost (one-off)	£46,413.11	**Error! Reference source not found.**
	Total (excluding HSCT)	£3,820.53	**Error! Reference source not found.**
Death	Event cost	£12,066.00	**Error! Reference source not found.**

Abbreviations: AML, acute myeloid leukaemia; EFS, event free survival; HCRU, healthcare resource utilisation; MDS, myelodysplastic syndrome; PSSRU, Personal Social Services Research Unit.

C13. Please provide a justification for the high wastage of Busulfan. Is it not possible to prepare (in an aseptic condition) a solution for multiple rounds of IV solution that can be stored in the fridge to prevent microbiological growth? In case this refers to any clinical guidelines (e.g. that state that Busulfan IV solutions always needs to be prepared freshly each time before administration), please provide the reference(s). In case this is based on clinical expert opinion, please provide the underlying argumentation.

The high wastage cost associated with busulfan is due to a combination of the dosing regimen received by patients in the trial (4 x 0.8 mg/kg/day at days -4 and -3 prior to alloHSCT), the average weight of patients in the phase III clinical trial (80.2 kg) and the single vial size option identified from the BNF (60 mg). When applying the dose in combination with the average patient weight, this generates a total dose size of 64.2 mg which is slightly above 60 mg vial available, resulting in remainder of 55.8 mg from the second vial administered. Due to the assumption of 100% wastage (based on clinical expert opinion) this means that for the base case analysis busulfan patients will be costed for 2 vials, at a total cost per dose of £382.38 (compared to £204.44 without wastage) and total treatment cost of £3,059 (compared £1,635,55 without wastage).

Overall wastage for treosulfan is lower for base case analysis due to the mean body surface area of patients in the phase III clinical trial (1.931 m^2) and the fewer dose administrations according to the dosing regimen from the clinical trial (1 x 10 g/m2/day at days -4, -3 and -2 prior to alloHSCT).

Variation of weight and body surface area was included in both the deterministic and probabilistic sensitivity analysis to allow for different wastage costs to be applied.

Given that the base case analysis generated a high wastage cost for busulfan, a scenario analysis was conducted where a single daily dose of 3.2 mg/kg was applied instead of 4 doses of 0.8 mg/kg/dayn. This dosing was investigated as the summary of product characteristics now allows a single dose but at the time of the pivotal trial it did not. This dosing regimen generated lower wastage costs due to a reduction in the vial remainder from the final vial administered (43.4 mg) and the smaller total number of separate drug administrations (2 instead of 8), resulting in a total dose cost of £817.77 (compared to £955.94 without wastage) and total treatment cost of £1,911.88 (compared £1,635,55 without wastage).

As shown in Scenario 5 in section B3.3.2.5 of the submission, this did not change the overall conclusions of the economic analysis.

The assumption of 100% wastage was based on consultation with the first clinical expert interviewed, who when asked “Do you anticipate that the remainder left over from Treosulfan and Busulfan vials when administering treatment will be recovered or that any excess will be wasted?” suggested assuming that the remainder would be wasted.

C14. The proportions of patients receiving either Flag/IDA or a second HSCT treatment (50:50) are based on clinical expert opinion.

Please justify, by providing the underlying argumentation, what these estimates are based on.

The proportions of patients receiving Flag/ISA or a second HSCT (50:50) were based on consultation with the first clinical expert interviewed. The clinical expert

was unsure of the exact proportion of patients that would receive but agreed upon suggestion from the interviewers that assumption of a 50/50 split was suitable in the absence of data. This assumption was also confirmed by the second clinical expert interviewed, who agreed it was a suitable assumption for the economic model.

Utilities

C15. Priority question: In section B3.3.2.4 it is stated, "However, two sets of disutilities for Grade 3 and 4 adverse events (AEs) were identified from TA399102 (0.0240 and 0.0207) based on different mappings of the QLQ-C30 data from the clinical trial, which were judged to be more aligned with the NICE reference case, despite concerns regarding potential differences in the AEs experienced by patients on a chemotherapy treatment (azacitidine) versus patients receiving HSCT".

Please further describe the likely differences between AEs experienced by patients on a chemotherapy treatment (azacitidine) versus patients receiving HSCT and discuss how the use of AE disutilities from the chemotherapy treatment may impact the model results.

The most common Grade 3 or 4 AEs (occurring in ≥10% of subjects in the azacitidine arm) reported in the AML-AZA-001 trial described in section 4.12.1 of the company submission in TA399 were febrile neutropenia, neutropenia, thrombocytopenia, pneumonia, AML and anaemia. The most common treatment related AEs (occurring in ≥10% of subjects in the azacitidine arm) reported in the AML-AZA-001 trial were nausea, neutropenia, thrombocytopenia, febrile neutropenia, vomiting, decreased appetite, constipation, injection site reaction, diarrhoea, pyrexia and injection site erythema. While there is some overlap between these adverse events and the Grade 3 or 4 treatment-related AEs included in the economic analysis (febrile neutropenia, nausea, vomiting, diarrhoea), there are also some treatment-related Grade 3 or 4 AEs that occurred in the pivotal phase III trial (MC-FludT.14/L Trial II) that may not have occurred in the AML-AZA-001 trial (such as oral mucositis and elevated GGT). It is unclear whether the disutility estimates

from TA399 may overestimate or underestimate the disutility associated with serious AEs experienced by patients undergoing alloHSCT.

However, the choice of adverse event disutilities for non-GvHD related Grade 3 or 4 AEs included in the economic model does not appear to have a large impact on the outcomes. When applying the larger disutility estimates of -0.218, -0.176, -0.060 and -0.100 from Stein 2018 for serious infections (applied to sepsis and lung infection events), severe diarrhoea, severe redness/skin peeling (applied to oral mucositis and maculopapular rash events) and abnormally low blood counts (applied to febrile neutropenia events), the total incremental QALYs for treosulfan decreases by 0.0000572.

C16. Priority question: In section B3.3.2.5, it is stated that "Two separate mapped EQ-5D estimates of 0.623 and 0.568 based on QLQ-C30 clinical trial data were available from TA399, generated using the Proskorovsky et al. 2014 and McKenzie et al. 2009 mapping studies respectively. While these values initially appeared to have improved face validity compared to alternative EQ-5D related estimates, the estimates were from an elderly patient population with a mean age of 75 years, which was significantly older than the treosulfan phase III trial population."

Please justify why this age difference is problematic, despite being similar to the age difference between the trial population and the Grulke source of average age 40-45. Please justify why this age difference could not be handled similarly to the Grulke source, by adjusting for age.

While the absolute age difference between the trial population (MC-FludT.14/L Trial II) and patients in the Grulke study was similar to the age difference between the trial population (MC-FludT.14/L Trial II) and patients in the AML-AZA-001 trial (TA399), the age differences are in opposite directions, which has an important impact on the age adjustment calculations using the regression model from Ara and Brazier.

The utility function multipliers, applied to the full general population curve to estimate the age and sex adjusted functions for each health state, are estimated by dividing the raw utility estimate by the general population utility from the Ara and Brazier

regression model for the average age of the patients from the original study. As the general population utility decreases with increasing age, this produces smaller general population reference utilities (the denominator of the utility function multiplier calculation) for studies with older patient populations.

The raw utility estimates from the Grulke study for patients at discharge and for relapse patients from TA399 (mapped using Proskorovsky) were 0.660 and 0.623 respectively. However, when matching to the average age of patients from the original studies (~42.5 and 75 years respectively), the reference utilities from the general population function were 0.894 for the Grulke study and 0.758 for TA399, which generated a general population function utility multiplier for the Grulke discharge health state of 0.738 compared to 0.823 for the relapse health state from TA399, which was much closer to the ≤6 months short term recovery health state multiplier from Grulke (0.845). This lacked face validity given the expectation that patients relapsing following an HSCT would have lower health state utilities than patients in recovery who are discharged from hospital, an assumption that was validated with both of the clinical experts interviewed. This assumption was also consistent with the poor survival prognosis for patients relapsing following an HSCT compared to event-free patients observed in the phase III clinical trial. The second clinical expert interviewed also suggested that the health state utility for relapse patients might be lower than the relapse/progression utility used in the base case analysis, which provided further confirmation for our assumption that relapse patients would be expected to have lower utilities than event free patients in recovery following an HSCT.

C17. Priority question: In light of so many HRQoL options being identified, please conduct a meta-analysis of utility values for the following groups of identified utility values: EQ-5D direct, EQ-5D mapped, DCE/TTO.

The NICE reference case does not explicitly recommend conducting meta-analyses on utility data. As stated in Section 5.3.8, “when more than 1 plausible set of EQ-5D data is available, sensitivity analyses should be carried out to show the impact of the alternative utility values.” Section 5.3.9 also states that “sensitivity analyses to

explore variation in the use of the mapping algorithms on the outputs should be presented."

We believe that the scenario analyses of utility data presented in the original submission address these recommendations and that the choice of base case utility values is sufficiently justified from critical review of the studies identified and clinical expert opinion. As indicated in Section B3.6.3 of the submission, the use of alternative utility estimates explored in scenario analysis did not significantly impact the overall outcomes of the analysis.

We do not believe that conducting a meta-analysis of utility estimates identified for the submission is necessary or appropriate, as it would be subject to considerable heterogeneity.

As described in Peasgood and Brazier,[28] "utility scores using tariffs from other countries reflect different sets of preferences, and unless it is believed that preferences should be universal, or the value sets are very similar, the rational for pooling utilities that use different country-specific tariffs is not clear. Considerable intercountry differences in the social tariff of the EQ-5D have been identified, with differences varying across the EQ-5D distribution". From the systematic and targeted literature reviews conducted for the submission, no sources of direct EQ-5D data derived using UK EQ-5D tariffs were available, with two studies (Leunis 2014,[29] Uyl de Groot 1998[30]) using Dutch EQ-5D tariffs, three studies (Kurosawa 2016[31], Kurosawa 2015,[32] Kurosawa 2014[33]) using Japanese EQ-5D tariffs and one study (Mamolo 2017[34]) using US EQ-5D tariffs. Cross comparison of EQ-5D index population norms from Szende et al suggest differences among country-specific TTO value sets for the Netherlands, Japan and the US, as well as the UK, particularly among older patient groups.[35] For example, for the 55-64 age bracket, which most closely aligns with the average age of patients at baseline (59.6 years) in the pivotal phase III trial (MC-FludT.14/L Trial II), the average EQ-5D index populations norms for the Netherlands, US and Japan are 0.886, 0.830 and 0.881 (compared to the UK estimate of 0.799) suggesting considerable differences between populations.

Peasgood and Brazier also highlight that "the pooling of utility values should only be attempted where the data are valuing the same clinical health state for the

appropriate population". This is relevant for the economic analysis presented in the submission as it is not clear whether all the studies describe the same clinical health states, even when considering subgroups of studies that share the same EQ-5D tariff. For the two Dutch EQ-5D tariff studies, Leunis 2014[29] uses "no relapse" and "relapse" as health state descriptors, whilst Uyl de Groot 1998[30] uses "complete response" and "no complete response". Additionally, Uyl de Groot describes separate time-based health states (6 months, 1 year, 2 years) within the complete response and no complete response groups rather than providing an overall estimate of complete response and no complete response patients, whereas Leunis provides single utility estimates for no relapse and relapse patients. There also appear to be key differences in terms of baseline age between these studies, with a mean patient age of 52.7 years in the Leunis 2014 study, and a median age of 68 years in the Uyl de Groot study, in which only patients above the age of 60 were included. As with a conventional meta-analysis of clinical outcomes, minimising patient heterogeneity between studies is key to producing valid estimates.

Furthermore, it is important to note that in regard to the Japanese EQ-5D tariff studies, Kurosawa 2016[31] and Kurosawa 2015[32] are studies of the same patient population and were included as separate options in the economic analysis as slightly different EQ-5D estimates were provided in each study. In addition, it is unclear whether the EQ-5D outcomes from the Kurosawa 2014[33] abstract are from a distinct patient population compared to the other Kurosawa studies, and no estimates of uncertainty are provided for the EQ-5D data in this study to inform a meta-analysis.

As such, we believe that performing a meta-analysis of all directly elicited EQ-5D data would not be appropriate due to between-study heterogeneity and would be of limited value for an analysis from a UK healthcare perspective.

For mapped EQ-5D data, two sources were identified for inclusion in the economic analysis (TA399,[36] Grulke 2012[37]). While both sources provide estimates of quality of life during HSCT, there are considerable differences in the post-HSCT health state descriptions, with TA399 including one estimate of post-HSCT quality of life compared to the five estimates from Grulke 2012 at various time points following HSCT. Furthermore, the underlying QLQ-C30 data used to estimate the mapped

EQ-5D estimates were not available from TA399 and no estimates of precision (such as standard error, 95% CI, etc.) were provided for the mapped EQ-5D data. As such, the TA399 EQ-5D estimates would be subject to assumptions around the precision of the estimates which would likely impact any outcomes from a meta-analysis (particularly given the availability of only two studies).

Therefore, we also believe that a meta-analysis of mapped EQ-5D data is not appropriate.

Regarding the six TTO/DCE based utility estimates, two studies (Szende 2009,[38] Goss 2006[39]) were conducted on an MDS patient population rather than an AML population (as in the other four TTO/DCE studies) and provide utility estimates for different clinical health state descriptions based on transfusion dependence. For the two TTO studies focusing on an AML patient population (Castejón 2018,[40] Joshi 2019[41]), different health state descriptions were used for the health states most relevant to the economic model, with Joshi 2019 including health states specific to post-HSCT recovery patients and a combined treatment failure/relapse/refractory health state (compared to distinct relapse and refractory states in Castejón 2018). As highlighted above, Peasgood and Brazier recommend that “the pooling of utility values should only be attempted where the data are valuing the same clinical health state for the appropriate population”.

It is also important to note that the utility estimates are not bounded consistently across the TTO/DCE studies, with four studies bounded between 0 and 1 (Szende 2009,[38] Goss 2006,[39] Stein 2018,[42] Stein 2019[43]) and two studies (Castejón 2018,[40] Joshi 2019[41]) bounded between -1 and 1. In addition, two studies (Stein 2018,[42] Stein 2019[43]) were conducted on the same patient population and were included as separate sources in the economic model due to slight differences in some of the health state utilities provided, and as such it would not be appropriate to include both studies in a meta-analysis.

Generally speaking, it is unlikely that health state vignettes provided to individuals in each study would include the same health state definitions even if the overall health state descriptions (such as “post-HSCT” and “relapse”) were consistent across studies, which may limit the ability to compare utilities across custom TTO/DCE

studies and, as indicated in Peasgood and Brazier, introduce another layer of uncertainty.

Considerable variability is also likely to be introduced when comparing across different elicitation methods (TTO and DCE) and across scales with different boundings. As noted in Peasgood and Brazier, “even where we have included only utility values on the same clinical health state, the identified utility values are still likely to show variability across instruments and elicitation methods”.

Therefore, we do not believe that a meta-analysis of the TTO/DCE utility estimates would be appropriate and would be subject to considerable heterogeneity.

C18. In section B3.3.2.5 of document B of the CS, when discussing the possibility of using relapse utility values from Leunis or Mamolo, it is stated “However, in both cases, the estimates lacked face validity in terms of application to post-HSCT relapse patients, particularly when adjusted for age and sex, with health state utilities generated similar to (or in some cases exceeding) short-term post-HSCT remission estimates.

Please provide the age and sex adjusted utilities from these two sources in comparison to the base-case alternative utility values selected and justify why relapse utility values cannot be “similar to” short-term post-HSCT remission estimates.

The general population utility multipliers for the relapse health state generated from Leunis and Mamolo were 0.906 and 0.885 respectively. These general population utility function multipliers are higher than both the discharge utility multiplier (0.738) and ≤6 months short term recovery health state multiplier (0.845) from the Grulke study and were more similar to the 7-12 month post-HSCT recovery multiplier (0.915). For the baseline age of 59.6 years, the age and sex adjusted utilities for Leunis and Mamolo were 0.755 and 0.738, compared to 0.615, 0.704 and 0.762 for discharge, ≤6 months post-HSCT recovery and 7-12 months post-HSCT recovery, respectively.

We agree that the original description that “in both cases, the estimates lacked face validity in terms of application to post-HSCT relapse patients, particularly when adjusted for age and sex, with health state utilities generated similar to (or in some cases exceeding) short-term post-HSCT remission estimates” lacked clarity. This was intended to be in specific reference to the ≤6 months post-HSCT recovery and 7-12 month post-HSCT recovery health states, rather than the discharge health state. In fact, for both Leunis and Mamolo, the relapse utilities both exceeded the discharge and ≤6 months post-HSCT recovery utilities following age and sex adjustment and were more similar to the 7-12 month post-HSCT recovery health state. These were therefore judged to lack face validity given the expectation that patients in a relapse health state would have a worse quality of life than patients in recovery following an HSCT, an assumption which was validated with the two clinical experts interviewed.

C19. In section B3.3.2.2 of Document B it is stated that the Messerer et al. 2008 study was excluded “due to concerns around the face validity of both the QLQ-C30 domain scores and mapped EQ-5D estimates for a long-term remission AML patient population”.

Please describe these face validity issues, providing relevant values and further justification for exclusion.

Mapped health state utilities for Messerer 2008 (unadjusted and adjusted),[44] based on the five mapping algorithms considered in the analysis, are described below in Table 2.

Table 2: Summary of utility estimates from Messerer 2008

	Proskorovsky 2014[45]	**Crott 2010**[46]	**McKenzie 2009**[47]	**Kontodimopoulos 2009**[48]	**Marriott 2016**[49]
Mapped EQ-5D utility estimates	0.729	0.648	0.418	0.789	0.663
Mapped EQ-5D utility estimates – age and sex adjusted†	0.691	0.614	0.396	0.748	0.628

Abbreviations: OLS, ordinary least squares; UK, United Kingdom

†Adjusted using baseline characteristics of the phase III trial (39.2% female, mean age 59.6 years) and the Ara and Brazier general population utility function

Patients in the Messerer study had a median age of 47 years and were in long-term remission after first-line treatment, with a relapse-free survival of at least 5 years. Across the different mapping algorithms, there was considerable variation in the mapped EQ-5D estimates, with the age and sex adjusted baseline utilities varying from 0.396 (McKenzie mapping) to 0.748 (Kontodimopoulos mapping). When comparing the age and sex adjusted values for long-term remission patients with mapped estimates from the Grulke study, the mapped estimates from Messerer were considerably lower for each mapping algorithm than the corresponding long-term recovery estimates from Grulke, as well as lower than both the ≤6 month and 7-12 month short-term recovery estimates. Similarly, when comparing the age and sex adjusted mapped estimates (based on the baseline patient demographics of the phase III trial) from Messerer using the Proskorovsky and Mckenzie algorithms to those from TA399, the remission utilities were also considerably lower (0.691 vs 0.833 and 0.396 vs 0.814), particularly for the Mckenzie mapped estimates.

Following review of the original QLQ-C30 data from the Messerer study, it was observed that the symptom scores for long-term remission patients were considerably higher than symptom scores for other remission or post-HSCT recovery health states reported in other studies collected during the SLR. For example, the mean pain symptom score for alloHSCT patients in the Messerer study was 71.8, which is considerably higher than the range of mean pain symptom scores reported in the Grulke (16-47) and Peric (17.46-47.50) studies, including the most severe health states from each respective study (during HSCT, post-HSCT with highly active cGvHD). Higher symptom score for the QLQ-C30 normally indicate higher symptom burden, and thus these symptom scores appeared implausible given that patients in the Messerer study were in long-term remission.

Validation

C20. Priority question: Please provide details about what validation efforts were performed in Section B.3.8 of the CS and the results of these validation efforts. Please note that Appendix L3 provides the questions submitted to the experts but not the answers. Furthermore, the Philips' checklist is a tool to assess the quality of cost effectiveness studies but it is not a tool for validation. Validation could be assessed for example (but not necessarily) with the help of the validation tool AdViSHE (https://advishe.wordpress.com/author/advishe/).

Validation work conducted with clinical experts is described in the response to question C2.

General logic and calculation checks were performed on the model by the health economists who developed it, along with another health economist who was not directly involved with the model development. Tests performed included logical checks of calculations, stress testing of parameters, and individual setting of inputs to zero to determine whether the expected outputs were generated.

Model implementation (electronic model)

C21. Priority question: Please answer the following questions regarding the sheet Treosulfan_Engine_Pooled.

a. **Please explain formulae and underlying assumptions in columns BT:DM.**

As described in section B3.4.2 of the submission, adverse event disutilities and costs were spread over time in order to apply more accurate discounting. Adverse events were assumed to occur at a constant rate within the time frame for which the event was captured in the clinical trial. Based on this assumption, adverse event probabilities were partitioned into "start cycle incidence", "intermediary cycle incidence" and "end cycle incidence" probabilities to take into account that the start

and end time frames in which the adverse events could occur would not align with a whole number of 28-day model cycles.

Mean duration estimates for each adverse event were used to spread disutilities and costs over multiple cycles to reflect that some adverse events would last for more than one 28-day cycle (and would not divide into a whole number of cycles). Separate cycle duration “weights” (start/intermediary/end) were then calculated in order to estimate the proportion of the total event disutility or cost that would be attributed within the start, intermediary and end cycles from which the event initially occurred.

For the model engines, a “cycle of event initiation” was used to separate patients according to when the event occurred. For example, for events occurring in the first possible cycle (cycle 4 for extensive chronic GvHD), the start cycle incidence was used to estimate the proportion of patients experiencing the event in that cycle. This was applied in conjunction with the duration weights for start cycle incidence patients to spread the disutility and cost of the event over multiple cycles (for events lasting more than 28 days). Events occurring within the remaining time frame prior to the last cycle in which the event could occur were assigned the intermediary cycle incidence probability (and associated start, intermediary and end cycle duration weights), with events occurring in the last cycle in which the event occur assigned the end cycle incidence probability (and associated start/intermediary/end cycle duration weights).

b. **Please explain formulae and underlying assumptions in columns FM:HF.**

Please see response to part (a).

c. **Please explain why Dead health state has an associated disutility (column DR).**

The death health state has an associated disutility which is derived from patients who experienced adverse events during the cycle in which they died.

d. **Please clarify why some negative costs appear at the end of the models time horizon (column FA).**

Regarding negative costs appearing at the end of the time horizon, this was an error in the calculation engine and has been corrected in the updated Excel model. Updated results have been added to the bottom of the document. Please note that the overall interpretations and conclusions of the economic evidence have not changed.

C22. Priority question: Please explain the calculations in sheet "Safety". In particular, those in row 96 and below.

As described in the response to question C21, incidence estimates were spread across multiple cycles according to the time frame in which data for the adverse event was collected in the phase III clinical trial, and assuming that the adverse events occurred at a constant rate.

In addition, 100-day OS values for each of the parametric survival curves were used to spread the incidence of extensive chronic GvHD given the definition of chronic GvHD adopted for the phase III clinical trial.

C23. Utilities_functions sheet; columns R:AB; calculation of cycle utilities: Please explain why all cells are divided by 364 to get daily utilities.

Utilities were divided by 364 in accordance with the assumption that 1 year is equal to 364 days, as described in section B3.2.2.2 of the submission.

C24. Please confirm that the utility decrement with age is correctly implemented.

Disutilities associated with adverse events were not adjusted for age. This has been amended in the updated version of the excel model.

Updated results are presented at the end of the document. Please note that the overall interpretations and conclusions of the economic evidence have not changed.

C25. Please confirm that the results of the subgroups analyses are obtained using the subgroup-specific patient characteristics and not the pooled characteristics as shown in the Population sheet.

The subgroup analyses use the pooled patient characteristics to minimise heterogeneity; the primary purpose of the separate AML and MDS analyses was as a form of internal validation; they should (and did) generate similar results (when totalled) to the pooled AML / MDS base case analysis.

C26. Please explain why in "Treosulfan_dosing" 1000mg vials are not included in the calculations. Explain also why the total dose for treosulfan is multiplied by BSA and for busulfan it is multiplied by weight.

The cost calculation for treosulfan was based on a 10 g/m²/d dose and an average BSA of 1.931 m² (= the mean BSA of the 551 patients included in the pivotal treosulfan study). The total dose per day for such a patient would be 19,310 mg.

Every pharmacist would therefore use 4 vials of 5 g treosulfan and not 3 vials of 5 g + 5 vials of 1 g for example, which would be less convenient with respect to reconstitution and furthermore not cost effective (drug costs £832 vs. £893).

Most cytotoxic agents are dosed according to body surface area. However, Busulfan has a very narrow therapeutic index with side effects such as sinusoidal obstruction syndrome (SOS) of the liver and mucositis as a result of high busulfan exposure or increased incidence of graft failure due to low exposure to busulfan. Therefore, controlling the patient's AUC by therapeutic drug monitoring (TDM) became mandatory. Based on PK studies performed by the originator company (Pierre Fabre), a weight-based regimen was developed for busulfan which best fits to reach the target plasma level of 900 to 1,500 µmol-min/L.

In contrast, treosulfan has a much broader therapeutic index; drug monitoring is not necessary and treosulfan can be dosed (as most cytotoxic agents) according to body surface area. A PK covariate analysis revealed that BSA was the only clinically relevant covariate for clearance and volumes of distribution of treosulfan.[50]

C27. Please explain the rationale for including mortality costs.

Mortality costs are included to reflect the additional resource use that patients would be anticipated to require prior to death. Inclusion of mortality costs is also consistent with two recent prior AML technology appraisals (TA523[51] and TA545[52]).

C28. Please explain the difference between the curves labelled as 'Selection' and 'All data' in EFS and OS input sheets. What was used in the model and why?

The "selection" nomenclature is an artefact from when different cuts of data were being considered early in the model development. It was not considered appropriate to use different cuts of data, so this function was removed – the data shown are identical for both.

C29. Deterministic sensitivity analyses are based on 20% variation, which seems arbitrary. Please justify this choice and use limits from confidence intervals where possible.

The objective of the deterministic sensitivity analysis was to identify the most sensitive inputs in the model and as such as a consistent proportional variation was applied to each input rather than on the basis of confidence interval sizes that would vary proportionally between inputs, and potentially provide misleading results given the covariance between survival function parameters.

The choice of 20% variation is a relatively standard approach in cost-effectiveness analysis and is consistent with deterministic sensitivity analysis performed in two recent AML appraisals (TA399,[36] TA523[51]).

Section D: Textual clarification and additional points

Missing documents

D1. Priority question: Please provide all additional tables, figures and appendices relating to the CSRs included in the submission (MC-FludT.14/L, MC-FludT.6/L, MC-FludT.7/AML, MC-FludT.8/MDS and MC-FludT.17/M).

Additional tables, figures and appendices relating to the previously submitted CSRs have now been provided.

D2. Section B.2.2.4 of the CS provides a brief description of a study comparing data from the European Society for Blood and Marrow Transplantation registry to the fludarabine/treosulfan arm of the MC-FludT.14/L study. This registry study is referenced as: European Society for Blood and Marrow Transplantation (EBMT), Iacobelli, S., Koster, L. & Biezen, A. Van. Re-analysis of EBMT-registry data on Fludarabine/Melphalan and Busulfan/Cyclophosphamide based conditioning treatment compared to Fludarabine/Treosulfan based conditioning of adult AML and MDS patients treated in MC-FludT.14/L phase III by matched pairs (2019).

This document appears to be missing from the submission. Please provide copies of the protocol and any study reports for this study.

The EBMT report[3] was excluded in error and the full EBMT report is now provided in PDF format along with our response. Given the nature of this registry report, there are no further appendices or protocol documents available.

D3. Appendix D of the CS: Please confirm that the study described in section B.1.5.1.2 is the key trial MC-FludT.14/L Trial II and not MC-FludT.16, as indicated in the section heading.

We confirm that the study described is the pivotal MC-FludT.14/L Trial II and not MC-FludT.16 which is a typographical error.

D4. The CS (document B) uses two names for the key trial (MC-FludT.14/L and MC-FludT.14/L Trial II). A third name (MC-FludT.14/L Trial) is used elsewhere, e.g. Table 5 of document A of the CS.

Please clarify the use of these names, e.g. do they refer to different parts of the same study and how do they relate to the CSRs provided?

Document B provides data for the key trial MC-FludT.14/L Trial II only and alternative forms of the name are typographical errors. Table 5 of document A should read "MC-FludT.14/L Trial II".

Other questions

D5. Figures 12 to 15, 17 and 18 require more explanation. It is unclear what selected model S means. Some discussion on the curves might also be provided (e.g. Figure 15 seems to underestimate OS for busulfan).

Selected model S refers to the chosen parametric survival curve described in the text – in the case of Figures 12 to 15, 17 and 18, this specifically refers to the NMCM lognormal models that were chosen for the base case analyses.

As discussed in section B3.3.1.3 of the submission, NMCM lognormal curves were chosen for the base case analysis due consistency of outcome in terms of AIC results. MCM lognormal curves also produced consistent AIC outcomes across all, but AIC statistics were generally slightly lower for the NMCM lognormal curves and thus these models were selected, given that there were no clear or consistent improvements in visual fit across the OS and EFS curves for both treatments.

Updated results

Updated results from Section B3.5, B3.6 and B3.7 are presented below, following amendments and corrections identified from C21d and C24. Please note that the changes to the model have not had a significant material impact on the results of the economic analysis with relatively minor changes in incremental costs and QALYs,

and as such the overall interpretations and conclusions of the economic evidence have not changed.

B.1.1 Base-case results

B.1.1.1 Base-case incremental cost-effectiveness analysis results

B.1.1.1.1 Base case deterministic results

The base-case deterministic results of the analysis are presented below in Table 3. Treosulfan was dominant over busulfan, with lower total costs and higher total QALYs.

Table 3: Base-case cost-effectiveness results

Technologies	Total costs (£)	Total LYG	Total QALYs	Incremental costs (£)	Incremental LYG	Incremental QALYs	ICER versus baseline (£/QALY)	ICER incremental (£/QALY)
Treosulfan	£140,308	8.73	6.44					
Busulfan	£163,976	7.71	5.55	£23,668	-1.01	-0.89	Dominated	Dominated

Abbreviations: ICER, incremental cost-effectiveness ratio; LYG, life years gained; QALYs, quality-adjusted life-years

B.1.1.1.2 Results by health state

Disaggregated results by health state are presented below in Table 4.

Table 4: Disaggregated results by health state

Health state	Costs		QALYs		LYG (undiscounted)	
	Treosulfan	Busulfan	Treosulfan	Busulfan	Treosulfan	Busulfan
Event-free survival	£102,989.86	£100,798.67	6.137	4.967	11.323	9.132
Relapse / progression: AML	£17,972.15	£33,816.13	0.193	0.372	0.476	0.965
Relapse / progression: MDS	£10,640.85	£20,049.21	0.109	0.210	0.269	0.546
Dead	£8,705.45	£9,312.29	-0.004	-0.007	12.069	10.643
Total	£140,308.31	£163,976.30	6.436	5.549	8.726	7.711

Abbreviations: AML, acute myeloid leukaemia; LYG, life year gain; MDS, myelodysplastic syndrome; QALYs, quality-adjusted life-year.

B.1.2 Sensitivity analyses

B.1.2.1 Probabilistic sensitivity analysis (PSA)

Probabilistic sensitivity analysis was performed with 5,000 simulations to capture stochastic uncertainty around each of the model inputs, with simultaneous variation for each parameter according to a predefined probability distribution. For inputs with no uncertainty data available, standard errors were estimated as 20% of the mean value. For all other inputs, standard errors were taken directly from the associated source, or derived from other measures presented (standard deviations, 95% CIs, etc.). A full list of inputs, associated standard errors and selected distributions are provided in Appendix L (model input list), as well as survival parameter values and Cholesky matrices for all survival functions and patients subgroups has been provided (Appendix M2, M3 and M4).

Results of the probabilistic sensitivity analysis for the base case model are shown in Figure 6 (cost-effectiveness plane) and Figure 7 (cost-effectiveness acceptability curve; CEAC).

Figure 6: Base case PSA results – cost-effectiveness plane

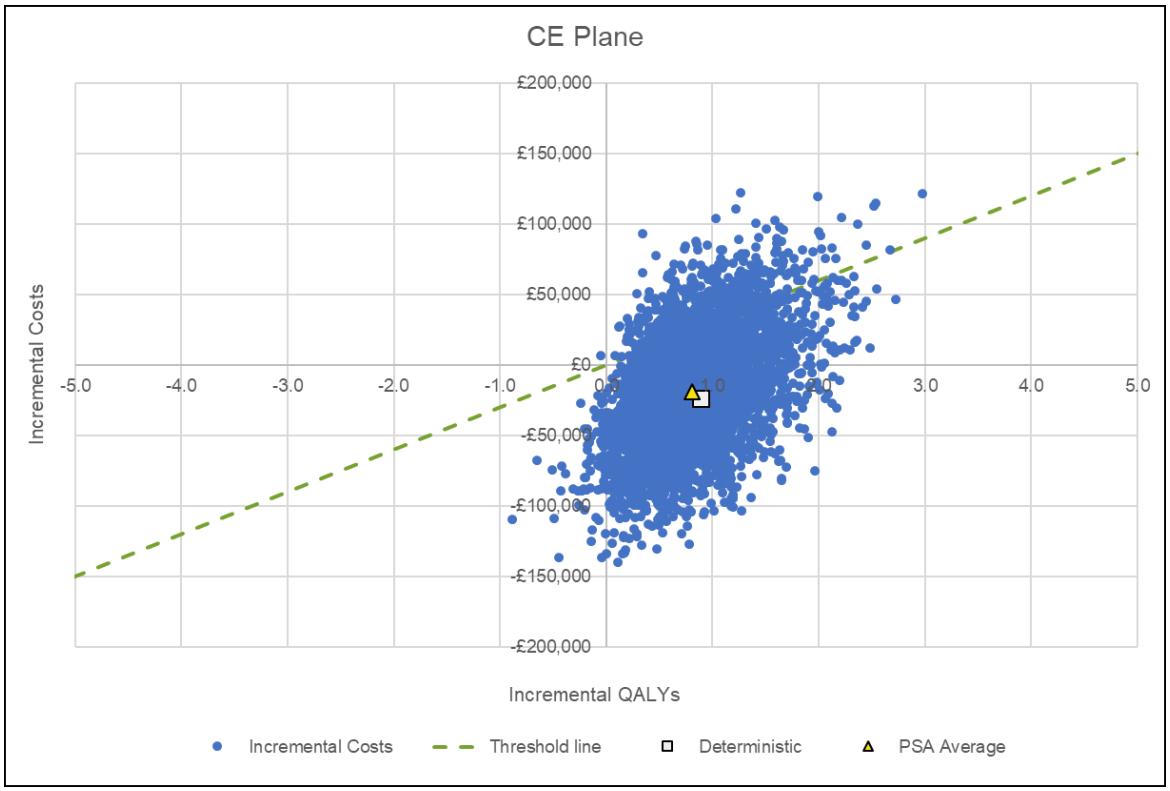
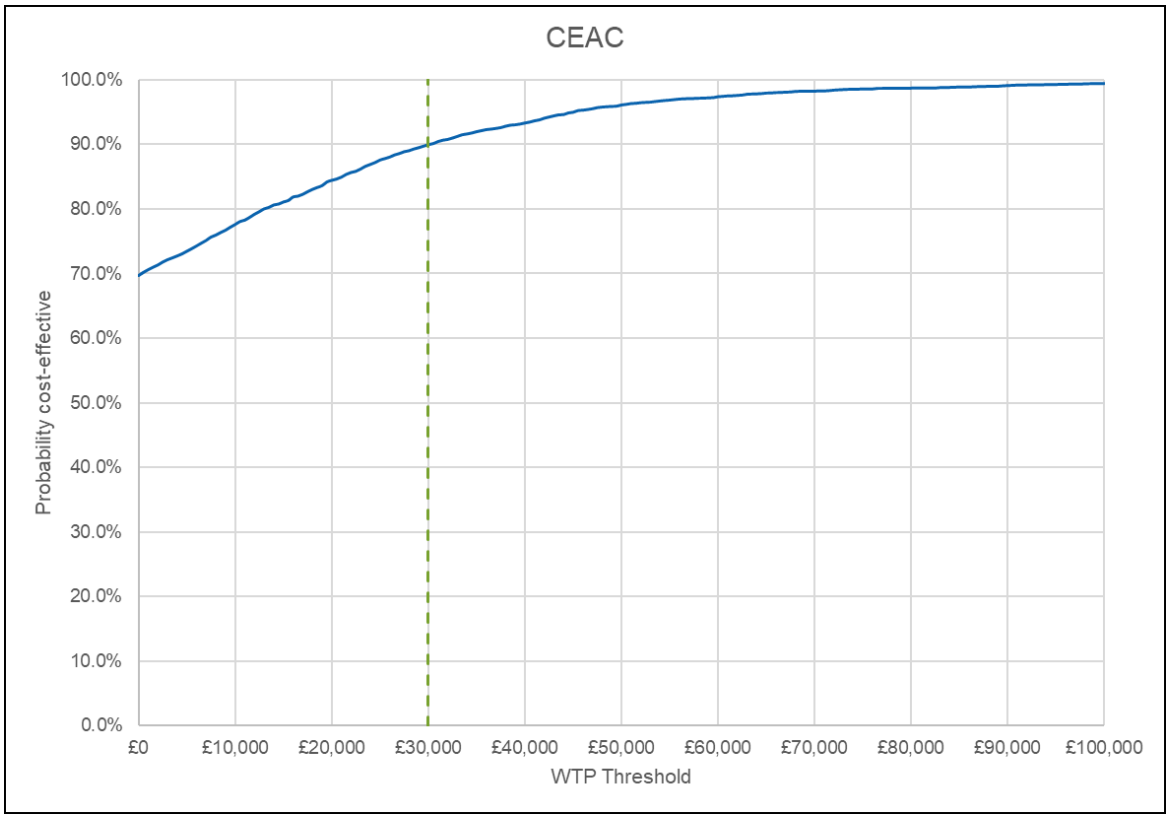


Figure 7: Base case PSA results – cost-effectiveness acceptability curve (CEAC)



The mean probabilistic incremental total costs were -£18,769 for treosulfan compared to busulfan, with mean probabilistic incremental total QALYs of 0.81. The mean incremental total life years were 1.10 for treosulfan compared to busulfan. The mean probabilistic ICER was -£23,312 (95% CI: -£19,821 to -£17,717 for incremental costs, 0.79-0.82 for incremental QALYs), with a mean INMB of £42,923 (based on a £30,000 per QALY threshold).

At a willingness to pay (WTP) threshold of £30,000 per QALY, the probability of treosulfan being cost-effective was 90.0%. Treosulfan was highly cost-effective at all thresholds examined, with a cost-effectiveness probability of 84.4% at a £20,000 per QALY and 69.7% probability at a £0 threshold.

B.1.2.2 Deterministic sensitivity analysis

Univariate sensitivity analysis was also performed, with each parameter varied by +/-20%. Results for the incremental cost-effectiveness ratio (ICER) are shown in Figure 8, with results for the incremental net monetary benefit (INMB) (using a £30,000 per QALY WTP threshold) shown in Figure 9.

The tornado diagrams indicate that the most sensitive inputs in the model were the meanlog coefficients for the NMCM lognormal survival functions for each treatment.

Figure 8: Base case DSA results – ICER

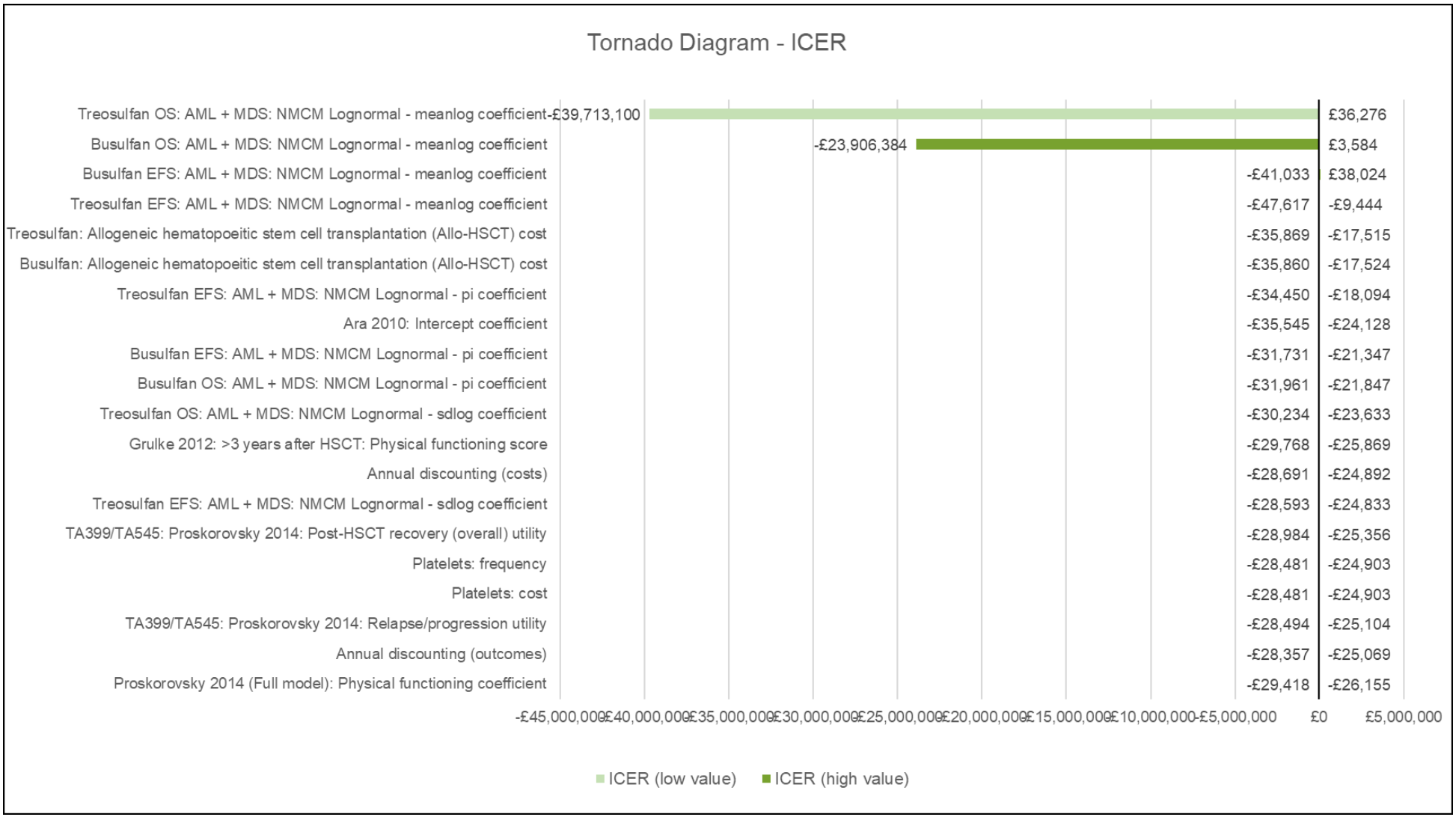
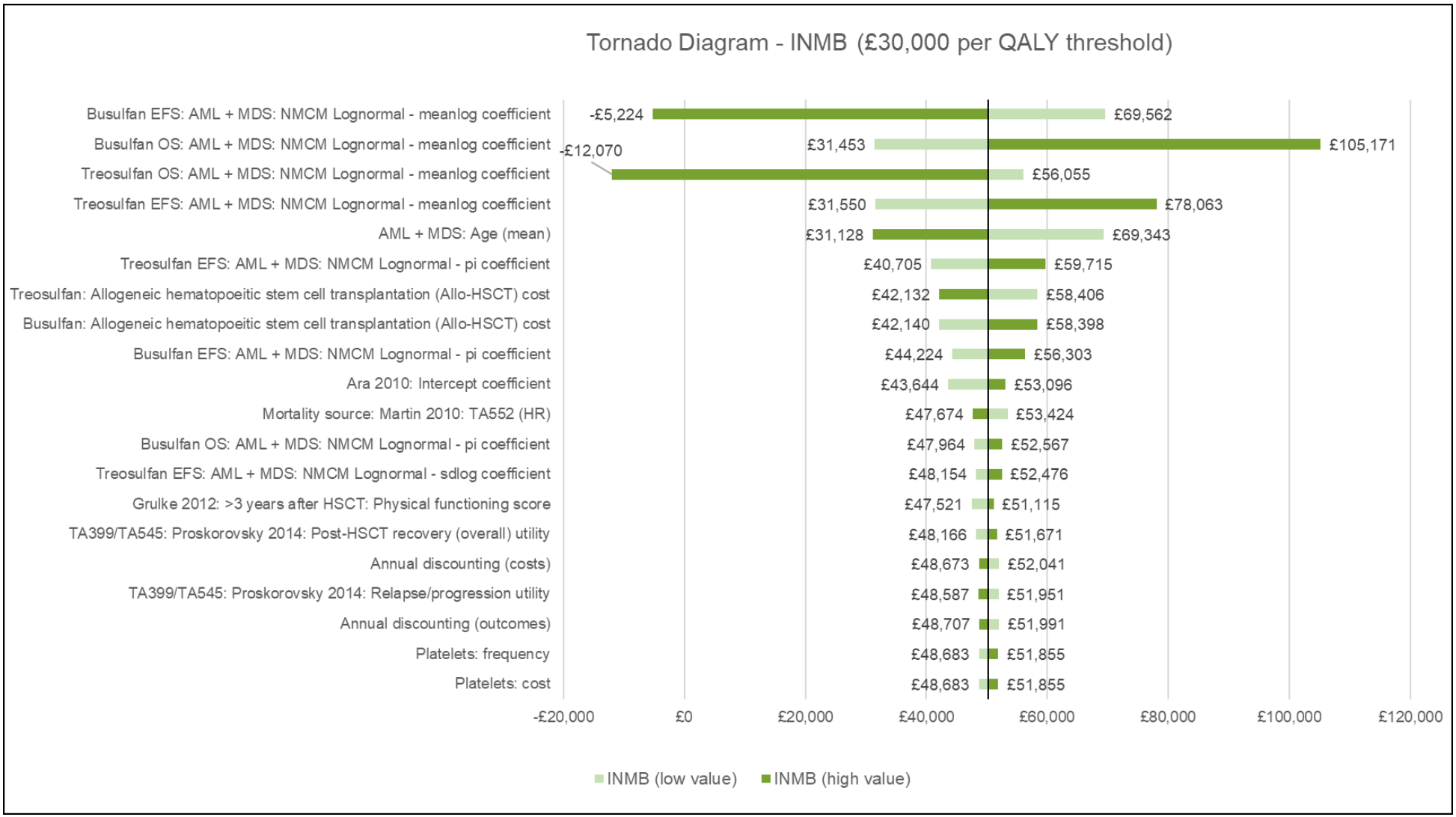


Figure 9: Base case DSA results – INMB (£30,000 per QALY threshold)



B.1.2.3 Scenario analysis

B.1.2.3.1 Scenario 1: Use of McKenzie et al 2009 mapping estimates

Three scenario analyses regarding utility values were performed. The first scenario was to utilise an alternative mapping (McKenzie et al 2009[47]) of the utility values estimated from Grulke et al 2012[37] and those used in TA399[36]. The age and sex adjusted utility parameters (for the baseline age of 59.6 years) are shown below in Table 5.

Table 5: Scenario 1: Use of McKenzie et al 2009 (for Grulke et al 2012 and TA399) utilities

Variable	Base case value (Proskorovsky et al 2014[45] mapping)	Updated scenario value (McKenzie et al 2009 mapping)[47]
Induction / HSCT utility	0.519	0.336
Post-HSCT recovery (short term) discharge	0.615	0.442
Post-HSCT recovery (short term) ≤6 months	0.704	0.569
Post-HSCT recovery (short term) 7-12 months	0.762	0.664
Post-HSCT recovery (long-term) year 2 and year 3	0.766	0.658
Post-HSCT recovery (long-term) year 4+	0.810	0.695
Functionally cured	0.810	0.695
Relapse / progression (AML and MDS patients)*†	0.569	0.436
Grade III+ AEs*	-0.026	-0.023

Source: McKenzie et al 2009,[47] Grulke et al 2012,[37] TA399,[36] and Proskorovsky et al 2014.[45]
Abbreviations: AE, adverse event; AML, acute myeloid leukaemia; HSCT, hematopoietic stem cell transplant; MDS, myelodysplastic syndrome
*Value reported directly in TA399.
†Estimated by applying a utility multiplier vs short-term post-HSCT recovery (≤6 months).

Deterministic results for the scenario analysis are shown below in Table 6.

Table 6: Scenario 1: Cost-effectiveness results

Technologies	Total costs (£)	Total LYG	Total QALYs	Incremental costs (£)	Incremental LYG	Incremental QALYs	ICER versus baseline (£/QALY)	ICER incremental (£/QALY)
Treosulfan	£140,308	8.73	5.47					
Busulfan	£163,976	7.71	4.68	£23,668	-1.01	-0.79	Dominated	Dominated

Abbreviations: ICER, incremental cost-effectiveness ratio; LYG, life years gained; QALYs, quality-adjusted life years

B.1.2.3.2 Scenario 2: Use of Castejon et al 2018 and Stein et al 2018 utility values

For the second scenario analysis, the base case HSUVs were replaced with those from Castejón et al 2018,[40] with serious AE disutilities utilised from Stein et al 2018.[42] As described in Section **Error! Reference source not found.**, the relapse/progression utilities were applied using the standard age-adjusted utility function, rather than as a disutility multiplier applied to short-term post-HSCT recovery patients (≤6 months) as per the base case model. A summary of the age-adjusted utilities for the baseline age and scenario analysis results are shown in Table 7 and Table 8 respectively.

Table 7: Scenario 2: Castejon et al 2018[40]/Stein et al 2018[42] utilities

Variable	Base case value	Updated scenario value (Castejon et al 2018[40]/Stein et al 2018[42])
Induction / HSCT utility	0.519	0.268
Post-HSCT recovery (short term) discharge	0.615	0.593
Post-HSCT recovery (short term) ≤6 months	0.704	0.593
Post-HSCT recovery (short term) 7-12 months	0.762	0.593
Post-HSCT recovery (long-term) year 2 and year 3	0.766	0.593
Post-HSCT recovery (long-term) year 4+	0.810	0.593
Functionally cured	0.810	0.726
Relapse / progression (AML and MDS)	0.569	0.096
Extensive chronic GvHD and stage III/IV acute GvHD disutility	-0.116	-0.182
Induction / HSCT utility	0.519	0.268
Sepsis/lung infection	-0.026	-0.204
Diarrhoea	-0.026	-0.165
Oral mucositis and maculopapular rash	-0.026	-0.056
Febrile neutropenia	-0.026	-0.094

Abbreviations: AML, acute myeloid leukaemia; GvHD, graft versus host disease; HSCT, hematopoietic stem cell transplant; MDS, myelodysplastic syndrome

Table 8: Scenario 2: Cost-effectiveness results

Technologies	Total costs (£)	Total LYG	Total QALYs	Incremental costs (£)	Incremental LYG	Incremental QALYs	ICER versus baseline (£/QALY)	ICER incremental (£/QALY)
Treosulfan	£140,308	8.73	5.29					
Busulfan	£163,976	7.71	4.33	£23,668	-1.01	-0.96	Dominated	Dominated

Abbreviations: ICER, incremental cost-effectiveness ratio; LYG, life years gained; QALYs, quality-adjusted life years

B.1.2.3.3 Scenario 3: Alternative MDS relapse/progression utility

For the third scenario analysis, an MDS specific utility for RBC transfusion dependent patients (from Szende et al 2009[38]) was applied to MDS relapse patients. Following application of the utility multiplier vs short-term HSCT-recovery (≤6 months), the age-adjusted utility for patients aged 59.6 years was 0.538; results from this analysis are shown below in Table 9.

Table 9: Scenario 3: Cost-effectiveness results

Technologies	Total costs (£)	Total LYG	Total QALYs	Incremental costs (£)	Incremental LYG	Incremental QALYs	ICER versus baseline (£/QALY)	ICER incremental (£/QALY)
Treosulfan	£140,308	8.73	6.43					
Busulfan	£163,976	7.71	5.54	£23,668	-1.01	-0.89	Dominated	Dominated

Abbreviations: ICER, incremental cost-effectiveness ratio; LYG, life years gained; QALYs, quality-adjusted life years

B.1.2.3.4 Scenario 4: Application of relapse utility multiplier to discharge utility

As described in Section **Error! Reference source not found.**, an additional utility scenario was explored where the relapse disutility multiplier from the Proskorovsky et al 2014[45] mapping of TA399 utilities was applied to discharge patients. Results are presented below in Table 10.

Table 10: Scenario 4: Cost-effectiveness results

Technologies	Total costs (£)	Total LYG	Total QALYs	Incremental costs (£)	Incremental LYG	Incremental QALYs	ICER versus baseline (£/QALY)	ICER incremental (£/QALY)
Treosulfan	£140,308	8.73	6.40					
Busulfan	£163,976	7.71	5.47	£23,668	-1.01	-0.92	Dominated	Dominated

Abbreviations: ICER, incremental cost-effectiveness ratio; LYG, life years gained; QALYs, quality-adjusted life years

B.1.2.3.5 Scenario 5: Alternative busulfan dosing

An additional scenario analysis was performed assuming a single 3.2mg/kg/day dose for busulfan (rather than 4 x 0.8mg/kg per day doses as per the trial population). Results are shown below in Table 11.

Table 11: Scenario 5: Cost-effectiveness results

Technologies	Total costs (£)	Total LYG	Total QALYs	Incremental costs (£)	Incremental LYG	Incremental QALYs	ICER versus baseline (£/QALY)	ICER incremental (£/QALY)
Treosulfan	£140,288	8.73	6.44					
Busulfan	£162,813	7.71	5.55	£22,524	-1.01	-0.89	Dominated	Dominated

Abbreviations: ICER, incremental cost-effectiveness ratio; LYG, life years gained; QALYs, quality-adjusted life years

B.1.2.3.6 Scenario 6: Separate modelling of AML and MDS

Another scenario was considered where AML and MDS patients were modelled separately in terms of EFS and OS, i.e. using AML and MDS-specific survival data rather than pooling the patients together to generate combined AML + MDS overall survival and event-free survival. To generate pooled results, the overall results for AML and MDS were weighted by the proportion of patients with AML and MDS pre-transplant (63.88% and 36.12% respectively).

The Kaplan-Meier plots and fitted models for each (NMCM lognormal in all instances) are shown below in Figure 10,

Figure 11, Figure 12, Figure 13, Figure 14, Figure 15, Figure 16, and Figure 17.

Figure 10: Kaplan-Meier versus survival model EFS: AML patients, treosulfan arm

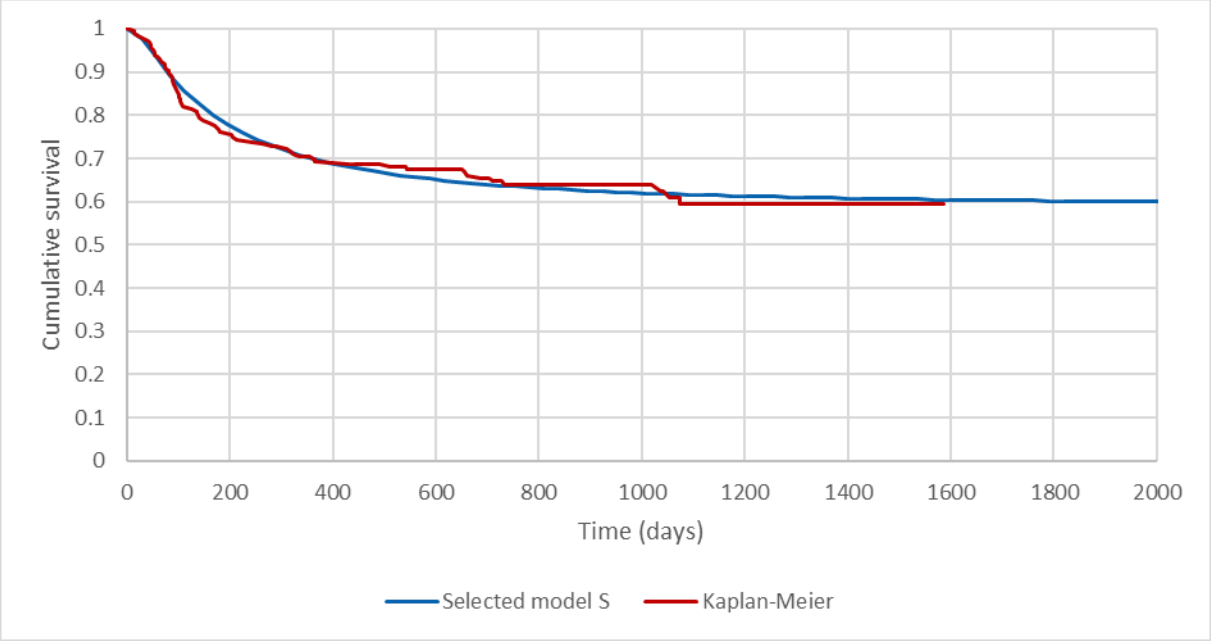


Figure 11: Kaplan-Meier versus survival model EFS: MDS patients, treosulfan arm

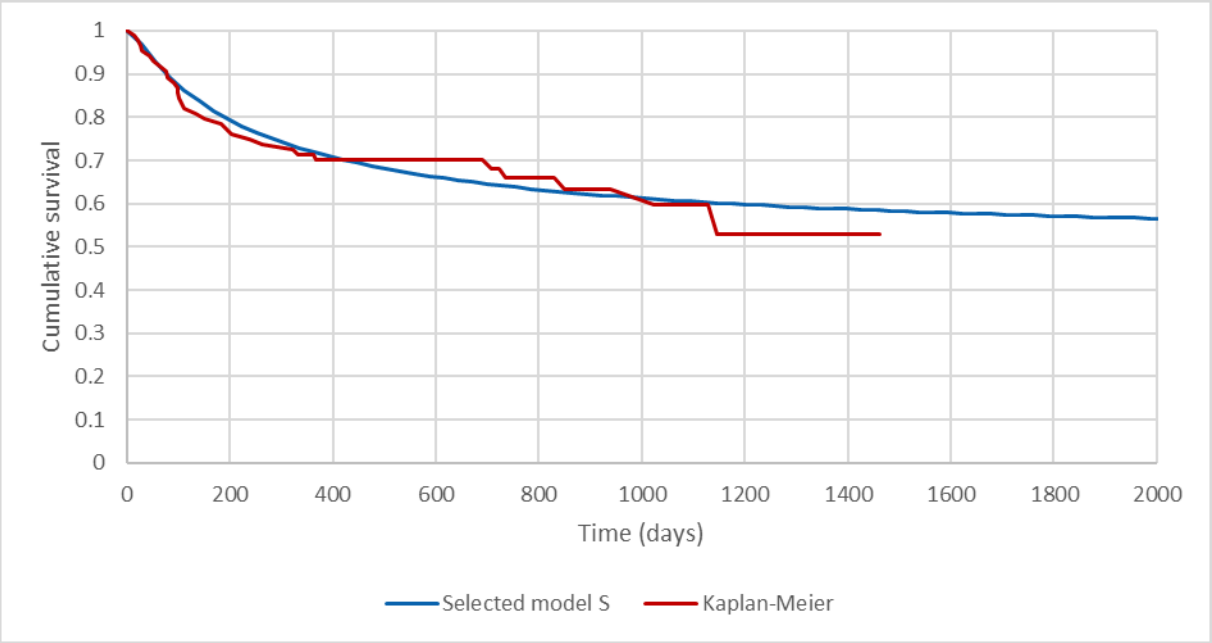


Figure 12: Kaplan-Meier versus survival model EFS: AML patients, busulfan arm

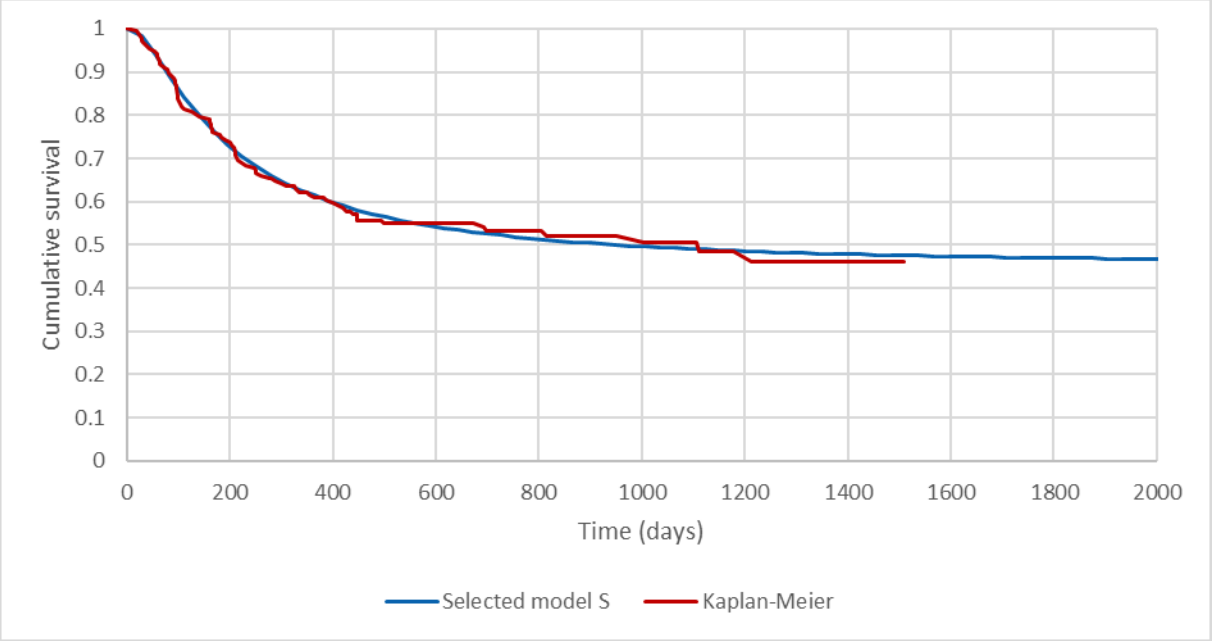


Figure 13: Kaplan-Meier versus survival model EFS: MDS patients, busulfan arm

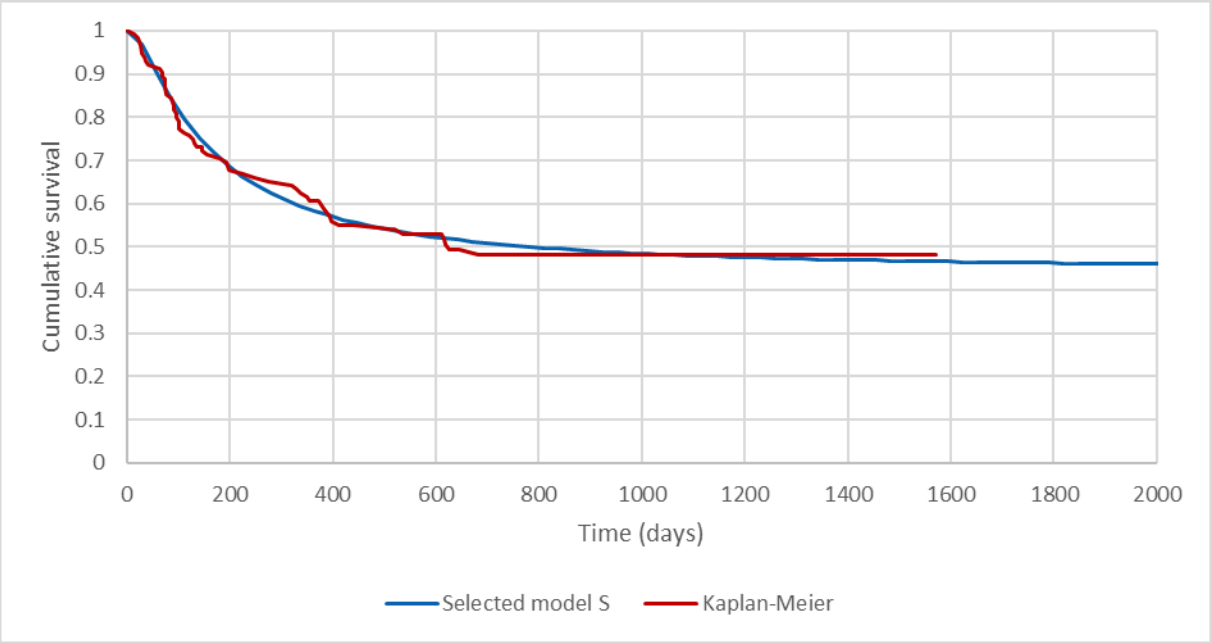


Figure 14: Kaplan-Meier versus survival model OS: AML patients, treosulfan arm

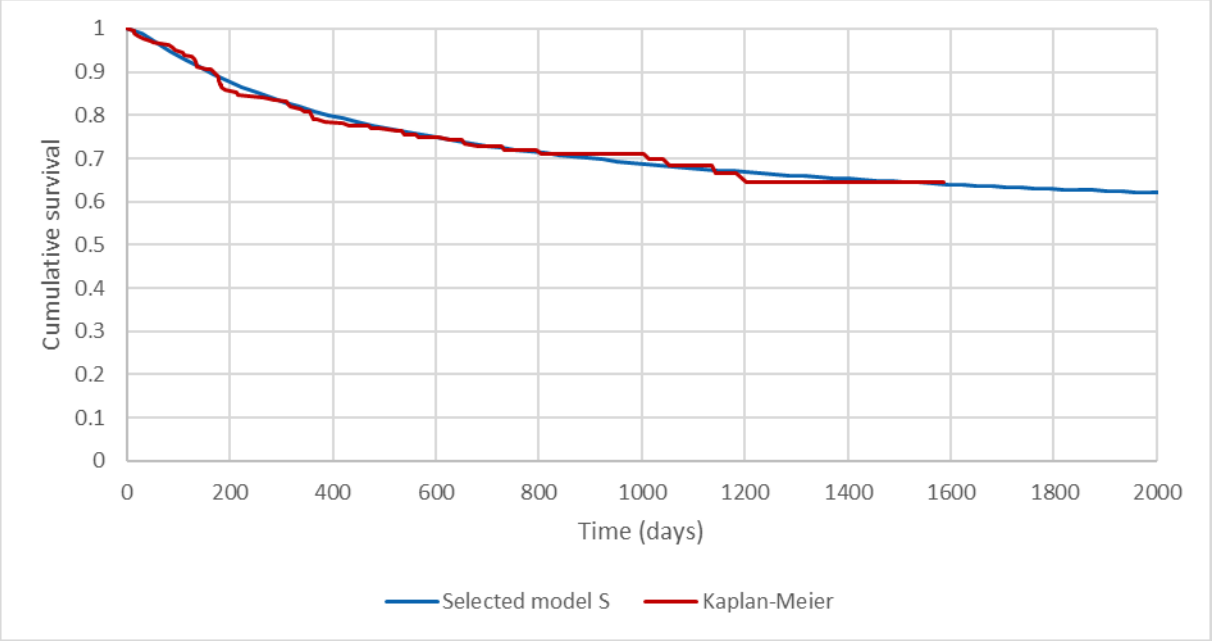


Figure 15: Kaplan-Meier versus survival model OS: MDS patients, treosulfan arm

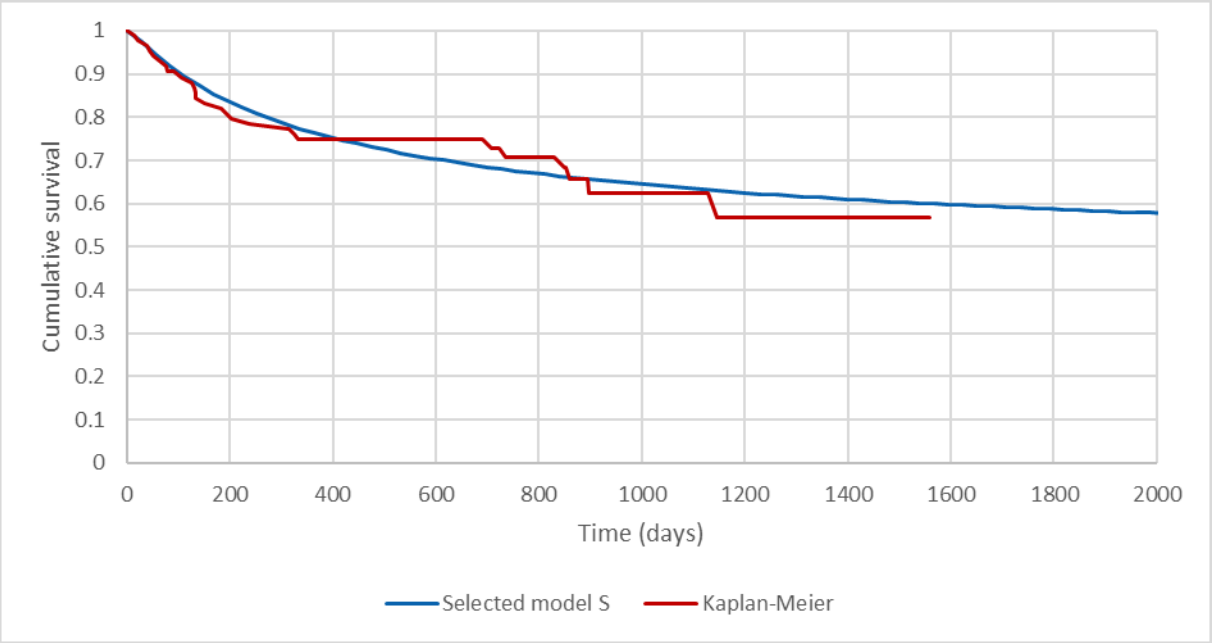


Figure 16: Kaplan-Meier versus survival model OS: AML patients, busulfan arm

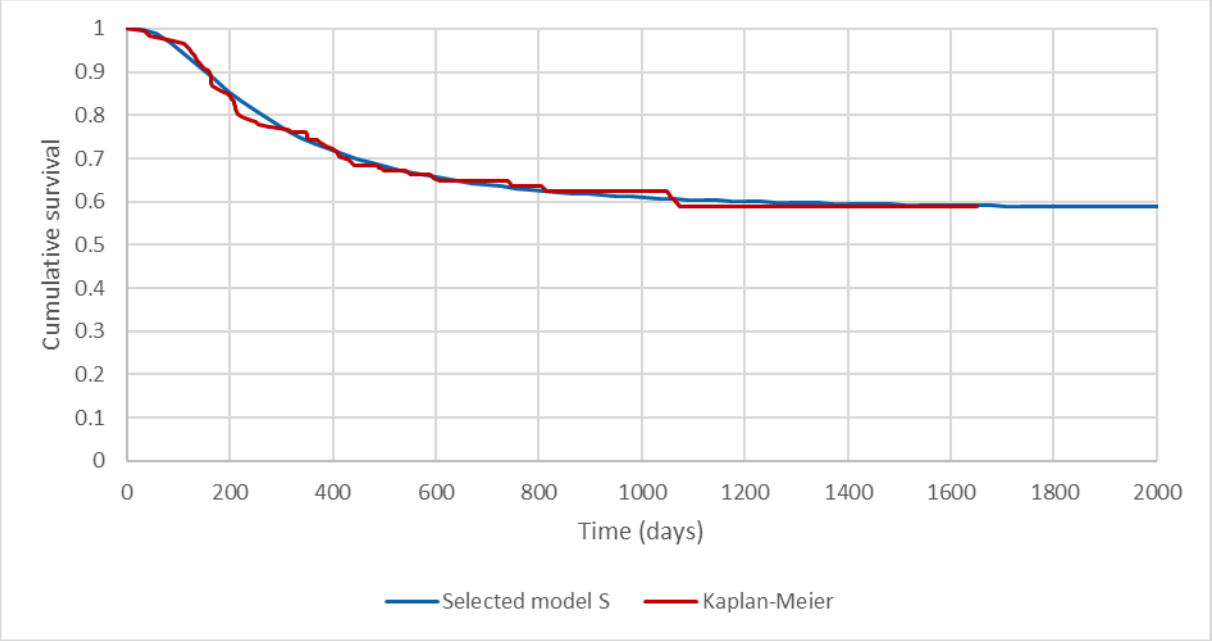
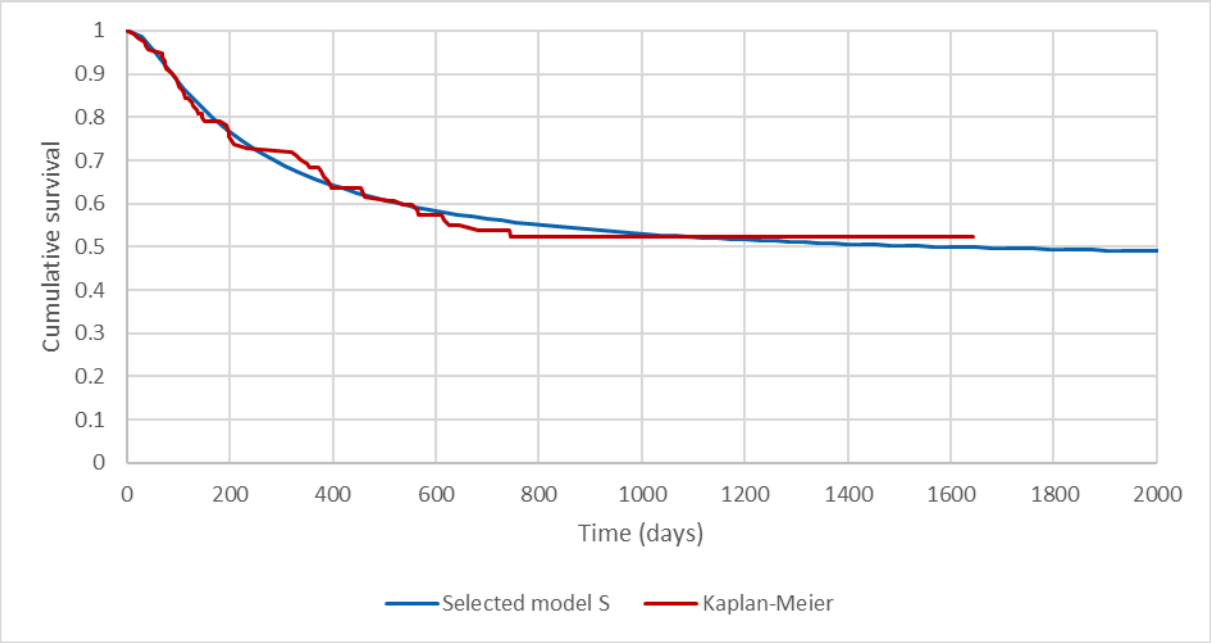


Figure 17: Kaplan-Meier versus survival model OS: MDS patients, busulfan arm



The results of this analysis are shown below in Table 12.

Table 12: Scenario 6: Cost-effectiveness results

Technologies	Total costs (£)	Total LYG	Total QALYs	Incremental costs (£)	Incremental LYG	Incremental QALYs	ICER versus baseline (£/QALY)	ICER incremental (£/QALY)
Treosulfan	£140,082	8.70	6.42					
Busulfan	£165,407	7.73	5.55	£25,324	-0.96	-0.87	Dominated	Dominated

Abbreviations: ICER, incremental cost-effectiveness ratio; LYG, life years gained; QALYs, quality-adjusted life years

B.1.2.3.7 Scenarios 7 and 8: Application of 5- and 10-year time horizons

Further scenarios were considered where 5- and 10-year time horizons were applied. Results are shown below in Table 13 and Table 14 respectively.

Table 13: Scenario 7: Cost-effectiveness results – 5-year time horizon

Technologies	Total costs (£)	Total LYG	Total QALYs	Incremental costs (£)	Incremental LYG	Incremental QALYs	ICER versus baseline (£/QALY)	ICER incremental (£/QALY)
Treosulfan	£118,941	3.32	2.44					
Busulfan	£123,995	2.98	2.15	£5,053	-0.34	-0.29	Dominated	Dominated

Abbreviations: ICER, incremental cost-effectiveness ratio; LYG, life years gained; QALYs, quality-adjusted life-years

Table 14: Scenario 8: Cost-effectiveness results – 10-year time horizon

Technologies	Total costs (£)	Total LYG	Total QALYs	Incremental costs (£)	Incremental LYG	Incremental QALYs	ICER versus baseline (£/QALY)	ICER incremental (£/QALY)
Treosulfan	£126,839	5.57	4.16					
Busulfan	£139,588	4.94	3.61	-£12,749	-0.63	-0.55	Dominated	Dominated

Abbreviations: ICER, incremental cost-effectiveness ratio; LYG, life years gained; QALYs, quality-adjusted life-years

B.1.2.3.8 Scenario 9: Application of gamma survival functions

As described in Section **Error! Reference source not found.** use of gamma survival functions was explored. Results of the scenario analysis are shown below in Table 15.

Table 15: Scenario 9: Cost-effectiveness results

Technologies	Total costs (£)	Total LYG	Total QALYs	Incremental costs (£)	Incremental LYG	Incremental QALYs	ICER versus baseline (£/QALY)	ICER incremental (£/QALY)
Treosulfan	£146,362	8.67	6.36					
Busulfan	£162,577	7.37	5.30	£16,215	-1.30	-1.06	Dominated	Dominated

Abbreviations: ICER, incremental cost-effectiveness ratio; LYG, life years gained; QALYs, quality-adjusted life-years

B.1.2.4 Summary of sensitivity analyses results

Probabilistic sensitivity analysis indicated that treosulfan was highly cost-effective compared to busulfan, with mean probabilistic incremental costs of -£18,769 and mean probabilistic incremental QALYs of 0.81, generating a mean ICER of -£23,312. Although these estimates differed slightly from the deterministic results (incremental costs of -£23,668, incremental QALYs of 0.89), the CEAC showed that treosulfan was still highly likely to be cost-effective at all thresholds considered, with a

probability of cost-effectiveness of 90.0% and 69.7% at WTP thresholds of £30,000 per QALY and £0 per QALY respectively. Increasing the number of simulations did not appear to significantly alter the difference in outcomes observed for the base case and probabilistic results.

A Cholesky decomposition is implemented in the analysis in order to reflect correlation between survival function parameters; however, this is not always sufficient to reduce the skew in the distribution of mean ICERs produced from random draws for the PSA.

Exclusion of the survival inputs from the PSA generated more similar mean probabilistic incremental costs to the base case analysis, albeit with limited impact on the mean probabilistic incremental QALYs. Exclusion of the survival function parameters also generated much smaller variance around costs and QALYs, and higher probabilities of cost-effectiveness for treosulfan at all thresholds considered.

The deterministic sensitivity analysis indicated that the most sensitive inputs in the model were the NMCM lognormal survival function meanlog parameters for treosulfan and busulfan. This was expected given the results from the PSA, and anticipation that independent variance of correlated parameters would potentially produce extreme values for the ICER.

With the exception of 20% increases for the meanlog coefficients for the treosulfan OS and busulfan EFS curves, which produced ICERs of £36,276 and £38,024 respectively, all parameter variations indicated that treosulfan was dominant over busulfan.

The majority of scenario analyses showed only relatively minor changes in total incremental costs and QALYs. Pooling of individual AML and MDS patient population results produced similar outcomes to the base case model.

As expected, shortening the model time horizon reduced the incremental costs and QALYs. Use of gamma survival functions reduced the difference in total costs by 31%, albeit with an increase of 20% in the difference in total QALYs. Regardless, in all scenario analyses tested, treosulfan dominated busulfan, with lower total costs and higher total QALYs.

B.1.3 Subgroup analysis

Subgroup analysis was performed for AML and MDS patients individually. As in scenario 5, NMCM lognormal functions were applied in for OS and EFS for both treatments and sub-populations.

Results for AML patients are shown in Table 16, with MDS patient results shown in Table 17.

Table 16: AML subgroup cost-effectiveness results

Technologies	Total costs (£)	Total LYG	Total QALYs	Incremental costs (£)	Incremental LYG	Incremental QALYs	ICER versus baseline (£/QALY)	ICER incremental (£/QALY)
Treosulfan	£146,004	8.90	6.54					
Busulfan	£181,908	8.13	5.77	£35,904	-0.76	-0.77	Dominated	Dominated

Abbreviations: ICER, incremental cost-effectiveness ratio; LYG, life years gained; QALYs, quality-adjusted life-years

Table 17: MDS subgroup cost-effectiveness results

Technologies	Total costs (£)	Total LYG	Total QALYs	Incremental costs (£)	Incremental LYG	Incremental QALYs	ICER versus baseline (£/QALY)	ICER incremental (£/QALY)
Treosulfan	£129,609	8.35	6.20					
Busulfan	£136,218	7.03	5.15	£6,610	-1.32	-1.05	Dominated	Dominated

Abbreviations: ICER, incremental cost-effectiveness ratio; LYG, life years gained; QALYs, quality-adjusted life-years

Both subgroup analyses indicated that treosulfan was dominant over busulfan. For the AML subgroup, cost savings for treosulfan were larger than the base case model, with smaller QALY gains. Conversely, the MDS subgroup generated smaller cost savings for treosulfan, albeit with larger QALY gains.

References

1. Down, J. D., Westerhof, G. R., Boudewijn, A., Setroikromo, R. & Ploemacher, R. E. Thiotepa improves allogeneic bone marrow engraftment without enhancing stem cell depletion in irradiated mice. *Bone Marrow Transplant.* **21**, 327–330 (1998).

2. Medac. *MC-FludT.14/L Trial II - Clinical Study Report*. (2018).

3. European Society for Blood and Marrow Transplantation (EBMT), Iacobelli, S., Koster, L. & Biezen, A. Van. *Re-analysis of EBMT-registry data on Fludarabine/Melphalan and Busulfan/Cyclophosphamide based conditioning treatment compared to Fludarabine/Treosulfan based conditioning of adult AML and MDS patients treated in MC-FludT.14/L phase III by matched pairs.* . (2019).

4. Center for International Blood and Marrow Transplant Research (CIBMTR). *Report of the comparison of conditioning regimens against medac data*. (2019).

5. Passweg, J. R. *et al.* Is the use of unrelated donor transplantation leveling off in Europe? The 2016 European Society for Blood and Marrow Transplant activity survey report. *Bone Marrow Transplant.* **53**, 1139–1148 (2018).

6. Beelen, D. W. *et al.* Final Results of a Prospective Randomized Multicenter Phase III Trial Comparing Treosulfan/Fludarabine to Reduced Intensity Conditioning with Busulfan/Fludarabine Prior to Allogeneic Hematopoietic Stem Cell Transplantation in Elderly or Comorbid Patients. *Blood* **130**, 521 LP – 521 (2017).

7. Beelen, D. *et al.* Final Evaluation of a Clinical Phase III Trial Comparing Treosulfan to Busulfan-Based Conditioning Therapy Prior to Allogeneic Hematopoietic Stem Cell Transplantation of Adult Acute Myeloid Leukemia and Myelodysplastic Syndrome Patients Ineligible to Stan. *Biol. Blood Marrow Transplant.* **25**, S3 (2019).

8. Casper, J. *et al.* Allogeneic hematopoietic stem-cell transplantation in patients

with hematologic malignancies after dose-escalated treosulfan/fludarabine conditioning. *J. Clin. Oncol.* **28**, 3344–3351 (2010).

9. Casper, J. *et al.* Allogeneic hematopoietic SCT in patients with AML following treosulfan/fludarabine conditioning. *Bone Marrow Transplant.* **47**, 1171–1177 (2012).

10. Ruutu, T. *et al.* Reduced-toxicity conditioning with treosulfan and fludarabine in allogeneic hematopoietic stem cell transplantation for myelodysplastic syndromes:Final results of an international prospective phase II trial. *Haematologica* **96**, 1344–1350 (2011).

11. Kröger, N. *et al.* Allogeneic stem cell transplantation after conditioning with treosulfan, etoposide and cyclophosphamide for patients with ALL: A phase II-study on behalf of the German Cooperative Transplant Study Group and ALL Study Group (GMALL). *Bone Marrow Transplant.* **50**, 1503–1507 (2015).

12. Deeg, H. J. *et al.* Transplant Conditioning with Treosulfan/Fludarabine with or without Total Body Irradiation: A Randomized Phase II Trial in Patients with Myelodysplastic Syndrome and Acute Myeloid Leukemia. *Biol. Blood Marrow Transplant.* (2018). doi:10.1016/j.bbmt.2017.12.785

13. Kalwak, K. *et al.* Prospective Clinical Phase II Results on Treosulfan-Based Conditioning Treatment of 70 Paediatric Patients with Haematological Malignancies. *Blood* **132**, 3354 (2018).

14. Nemecek, E. R. *et al.* Treosulfan, Fludarabine, and Low-Dose Total Body Irradiation for Children and Young Adults with Acute Myeloid Leukemia or Myelodysplastic Syndrome Undergoing Allogeneic Hematopoietic Cell Transplantation: Prospective Phase II Trial of the Pediatric Blood. *Biol. Blood Marrow Transplant.* **24**, 1651–1656 (2018).

15. Yerushalmi, R. *et al.* Fludarabine and treosulfan compared with other reduced-intensity conditioning regimens for allogeneic stem cell transplantation in patients with lymphoid malignancies. *Bone Marrow Transplant.* **50**, 1526–1535 (2015).

16. Shimoni, A. *et al.* Fludarabine and treosulfan conditioning is associated with a more favorable outcome after allogeneic stem cell transplantation in myelodysplastic syndrome. A survey on behalf of the Chronic Malignancies Working Party of EBMT.No Title. in *41st Annual Meeting of the EBMT* (2015).

17. Shimoni, A. *et al.* Intravenous Busulfan Compared with Treosulfan-Based Conditioning for Allogeneic Stem Cell Transplantation in Acute Myeloid Leukemia: A Study on Behalf of the Acute Leukemia Working Party of European Society for Blood and Marrow Transplantation. *Biol. Blood Marrow Transplant.* **24**, 751–757 (2018).

18. Saraceni, F. *et al.* Fludarabine-treosulfan compared to thiotepa-busulfan-fludarabine or FLAMSA as conditioning regimen for patients with primary refractory or relapsed acute myeloid leukemia: A study from the Acute Leukemia Working Party of the European Society for Blood and. *J. Hematol. Oncol.* **12**, 44 (2019).

19. Gran, C. *et al.* Treosulfan Conditioning for Allogeneic Transplantation in Multiple Myeloma Improved Overall Survival in Upfront Hematopoietic Stem Cell Transplantation — a Large Retrospective Study By the Chronic Malignancies Working Party of the EBMT. *Blood* **132**, 3464 (2018).

20. EWOG-MDS. Guidelines for Hematopoietic Stem Cell Transplantation (HSCT) in Childhood MDS and JMML for Patients enrolled in EWOG-MDS Studies. Version 1.3. in *EWOG-MDS Consensus Conference Freiburg* 1–19 (2016).

21. Faraci, M. *et al.* Gonadal function after Busulfan compared to Treosulfan in children and adolescents undergoing allogeneic hematopoietic stem cell transplantation. On Behalf of the Pediatric and Transplant Complications Working Parties of EBMT. *Biol. Blood Marrow Transplant.* **0**, (2019).

22. European Society for Blood and Marrow Transplantation (EBMT). *The EBMT Handbook Hematopoietic Stem Cell Transplantation and Cellular Therapies*. (2019).

23. British Society of Blood and Marrow Transplantation (BSBMT). BSBMT

Registry 2017. (2017). Available at: http://bsbmt.org/activity/2017/. (Accessed: 5th April 2019)

24. Socié, G. *et al.* Long-term survival and late deaths after allogeneic bone marrow transplantation. Late Effects Working Committee of the International Bone Marrow Transplant Registry. *N. Engl. J. Med.* **341**, 14–21 (1999).

25. Martin, P. J. *et al.* Life expectancy in patients surviving more than 5 years after hematopoietic cell transplantation. *J. Clin. Oncol.* **28**, 1011–1016 (2010).

26. Goldman, J. M. *et al.* Relapse and late mortality in 5-year survivors of myeloablative allogeneic hematopoietic cell transplantation for chronic myeloid leukemia in first chronic phase. *J. Clin. Oncol.* **28**, 1888–1895 (2010).

27. Gunnarsson, N. *et al.* Population-based assessment of chronic myeloid leukemia in Sweden: striking increase in survival and prevalence. *Eur. J. Haematol.* **97**, 387–392 (2016).

28. Peasgood, T. & Brazier, J. Is Meta-Analysis for Utility Values Appropriate Given the Potential Impact Different Elicitation Methods Have on Values? *Pharmacoeconomics* **33**, 1101–1105 (2015).

29. Leunis, A., Redekop, W. K., Uyl-de Groot, C. A. & Löwenberg, B. Impaired health-related quality of life in acute myeloid leukemia survivors: A single-center study. *Eur. J. Haematol.* **93**, 198–206 (2014).

30. Uyl-de Groot, C. A. *et al.* Cost-effectiveness and quality-of-life assessment of GM-CSF as an adjunct to intensive remission induction chemotherapy in elderly patients with acute myeloid leukemia. *Br. J. Haematol.* **100**, 629–36 (1998).

31. Kurosawa, S. *et al.* Decision Analysis of Postremission Therapy in Cytogenetically Intermediate-Risk Acute Myeloid Leukemia: The Impact of FLT3 Internal Tandem Duplication, Nucleophosmin, and CCAAT/Enhancer Binding Protein Alpha. *Biol. Blood Marrow Transplant.* **22**, 1125–1132 (2016).

32. Kurosawa, S. *et al.* Patient-reported quality of life after allogeneic

hematopoietic cell transplantation or chemotherapy for acute leukemia. *Bone Marrow Transplant.* **50**, 1241–1249 (2015).

33. Kurosawa, S. *et al.* Decision Analysis of Allogeneic Hematopoietic Stem Cell Transplantation Versus Chemotherapy in Cytogenetically Standard-Risk Acute Myeloid Leukemia in First Complete Remission: The Impact of FLT3-ITD Profile. *Blood* **124**, (2014).

34. Mamolo, C. M. *et al.* Cross-Sectional Survey of Symptoms and Health-Related Quality of Life of Adults with De Novo Acute Myeloid Leukemia (AML) in Clinical Practice. *Blood* **130**, 5660 LP – 5660 (2017).

35. Szende, A., Janssen, B. & Cabases, J. *Self-Reported Population Health: An International Perspective based on EQ-5D*. *Self-Reported Population Health: An International Perspective based on EQ-5D* (Springer, 2014). doi:10.1007/978-94-007-7596-1

36. National Institute for Health and Care Excellence (NICE). TA399: Azacitidine for treating acute myeloid leukaemia with more than 30% bone marrow blasts. *Technology Appraisal Guidance* (2016). Available at: https://www.nice.org.uk/guidance/ta399. (Accessed: 17th April 2019)

37. Grulke, N., Albani, C. & Bailer, H. Quality of life in patients before and after haematopoietic stem cell transplantation measured with the European Organization for Research and Treatment of Cancer (EORTC) Quality of Life Core Questionnaire QLQ-C30. *Bone Marrow Transplantation* **47**, 473–482 (2012).

38. Szende, A. *et al.* Valuation of transfusion-free living in MDS: results of health utility interviews with patients. *Health Qual. Life Outcomes* **7**, 81 (2009).

39. Goss, T. F. *et al.* Cost effectiveness of lenalidomide in the treatment of transfusion- dependent myelodysplastic syndromes in the United States. *Cancer Control* **13**, 17–25 (2006).

40. Castejón, N. *et al.* Social preferences for health states associated with acute myeloid leukemia for patients undergoing treatment in the United Kingdom.

Health Qual. Life Outcomes **16**, 66 (2018).

41. Joshi, N. *et al.* Health State Utilities for Acute Myeloid Leukaemia: A Time Trade-off Study. *Pharmacoeconomics* **37**, 85–92 (2019).

42. Stein, E. M. *et al.* Assessing utility values for treatment-related health states of acute myeloid leukemia in the United States. *Health Qual. Life Outcomes* **16**, 193 (2018).

43. Stein, E. *et al.* Cost Effectiveness of Midostaurin in the Treatment of Newly Diagnosed FLT3-Mutated Acute Myeloid Leukemia in the United States. *Pharmacoeconomics* **37**, 239–253 (2019).

44. Messerer, D. *et al.* Impact of different post-remission strategies on quality of life in patients with acute myeloid leukemia. *Haematologica* **93**, 826–833 (2008).

45. Proskorovsky, I. *et al.* Mapping EORTC QLQ-C30 and QLQ-MY20 to EQ-5D in patients with multiple myeloma. *Health Qual. Life Outcomes* **12**, 35 (2014).

46. Crott, R. & Briggs, A. Mapping the QLQ-C30 quality of life cancer questionnaire to EQ-5D patient preferences. *Eur. J. Heal. Econ.* **11**, 427–434 (2010).

47. McKenzie, L. & Van Der Pol, M. Mapping the EORTC QLQ C-30 onto the EQ-5D instrument: The potential to estimate QALYs without generic preference data. *Value Heal.* **12**, 167–171 (2009).

48. Kontodimopoulos, N., Aletras, V. H., Paliouras, D. & Niakas, D. Mapping the cancer-specific EORTC QLQ-C30 to the preference-based EQ-5D, SF-6D, and 15D instruments. *Value Heal.* **12**, 1151–1157 (2009).

49. Marriott, E. R., van Hazel, G., Gibbs, P. & Hatswell, A. J. Mapping EORTC-QLQ-C30 to EQ-5D-3L in patients with colorectal cancer. *J. Med. Econ.* **20**, 193–199 (2016).

50. Kinesis. *Final Report of the Population PK modeling*. (2011).

51. National Institute for Health and Care Excellence (NICE). TA523: Midostaurin for untreated acute myeloid leukaemia. *Technology Appraisal Guidance* (2018). Available at: https://www.nice.org.uk/guidance/ta523. (Accessed: 17th April 2019)

52. National Institute for Health and Care Excellence (NICE). TA545: Gemtuzumab ozogamicin for untreated acute myeloid leukaemia. *Technology Appraisal Guidance* (2018). Available at: https://www.nice.org.uk/guidance/ta545. (Accessed: 17th April 2019)

Appendix

Table S1. EmBase search strings for additional searches

Search	Parameter	Query	Items found (20 June 2019)
#1	Disease term	('acute myeloid leukemia' OR 'promyelocytic leukaemia' OR 'acute monocytic leukaemia' OR 'erythroleukemia' OR 'acute megakaryoblastic leukaemia' OR 'acute lymphoblastic leukaemia' OR 'myelodysplastic syndrome' OR 'multiple myeloma' OR 'acute myeloid leukaemia' OR 'acute myeloid leukemia' OR 'AML' OR 'acute nonlymphoblastic leukaemia' OR 'acute nonlymphoblastic leukaemia' OR 'hematologic malignancies' OR 'hematologic malignancy' OR 'haematologic malignancies' OR 'haematologic malignancy' OR 'Myeloid Malignancies' OR 'Lymphoid Malignancies' OR 'Lymphoid Malignancy' OR 'acute myelocytic leukaemia' OR 'acute myelocytic leukemia' OR 'acute myeloblastic leukaemia' OR 'acute myeloblastic leukaemia' OR 'acute myelomonocytic leukaemia' OR 'acute myelomonocytic leukemia' OR 'acute myelomonoblastic leukemia' OR 'acute monoblastic leukaemia' OR 'acute monoblastic leukemia' OR 'acute monocytic leukaemia' OR 'acute monocytic leukemia' OR 'acute erythroid leukaemia' OR 'Pure erythroid leukaemia' OR 'acute erythroblastic leukaemia' OR 'erythroleukaemia' OR 'acute megakaryoblastic leukaemia' OR 'acute lymphoid leukaemia' OR 'acute lymphoid leukemia' OR 'acute basophilic leukaemia' OR 'acute myelocytic leukaemia' OR 'acute myeloblastic leukemia' OR 'acute myelomonocytic leukaemia' OR 'acute myelomonoblastic leukaemia' OR 'acute monoblastic leukaemia' OR 'acute erythroid leukaemia' OR 'Pure erythroid leukaemia' OR 'acute erythroblastic leukaemia' OR 'erythroleukaemia' OR 'acute basophilic leukaemia' OR 'Myelodysplastic syndrome' OR 'Myelodysplastic syndromes' OR 'MDS' OR 'pre-leukaemia' OR 'pre-leukemia' OR 'smoldering leukaemia' OR 'Smoldering leukemia' OR 'Acute promyelocytic leukaemia' OR 'Acute promyelocytic leukemia' OR 'Acute lymphoblastic leukemia' OR 'Acute lymphocytic leukaemia' OR 'Acute lymphocytic leukemia' OR 'erythremia' OR 'Acute promyelocytic leukaemia' OR 'Acute lymphocytic leukaemia' OR 'polycythemia vera' OR 'vaquez osler disease' OR 'Natural killer cell lymphoblastic leukaemia' OR 'Natural killer cell lymphoblastic leukemia' OR 'Natural killer cell leukaemia' OR 'Natural killer cell leukemia' OR 'NK-cell lymphoblastic leukaemia' OR 'NK-cell lymphoblastic leukemia' OR 'NK-cell leukaemia' OR 'NK-cell leukemia' OR 'Acute NK-cell leukaemia' OR 'Acute NK cell leukemia' OR 'Acute Natural Killer cell leukaemia' OR 'Acute NK-cell leukaemia' OR 'Acute Natural Killer cell leukaemia' OR 'ANKL' OR	212,574

		'Multiple myeloma' OR 'B-lymphoblastic leukaemia' OR 'B-lymphoblastic leukemia' OR 'B lymphoblastic leukaemia' OR 'B lymphoblastic leukemia' OR 'B-ALL' OR 'T-lymphoblastic leukaemia' OR 'T-lymphoblastic leukemia' OR 'T lymphoblastic leukaemia' OR 'T lymphoblastic leukemia' OR 'T-ALL' OR 'APL' OR 'APML' OR 'Kahler* disease' OR 'Kahler disease' OR 'plasma cell myeloma' OR 'myelomatosis'):ti,ab	
#2	HSCT terms	(('allogenic' OR 'autologous') AND transplant*) OR (halpo* AND 'allograft') OR (('umbilical cord blood' OR 'peripheral stem cell' OR 'hematopoietic' OR 'Bone marrow' OR 'marrow') AND transplant*)):ti,ab	159,115
#3	HSCT terms	(transplant* OR 'allograft'):ti,ab	678,754
#4	All HSCT terms	#2 OR #3	678,754
#5	Disease Population	#1 AND #4	41,138
#6	Interventions	('total body irradiation' OR 'TBI' OR 'irradiation' OR 'clinical management' OR 'established clinical management' OR 'Best Supportive care' OR 'BSC' OR 'dose-reduced chemotherapy intensity' OR 'high-dose intensity' OR 'reduced-dose intensity' OR 'RIC' OR 'Reduced Intensity Conditioning' OR 'standard of care' OR 'SOC' OR 'preparative' OR 'Reduced Intensity'):ti,ab	398,994
#7	CEA Terms	('pharmacoeconomics' OR 'ICERs' OR 'Incremental Cost-Effectiveness Ratio' OR 'Cost-benefit analysis' OR 'Cost benefit analysis' OR 'cost effective analysis' OR 'cost-effective analysis' OR 'cost effectiveness ratio' OR 'cost-effectiveness ratio' OR 'cost of illness' OR 'cost efficiency analysis' OR 'cost-efficiency analysis' OR 'cost minimisation analysis' OR 'cost minimization analysis' OR 'cost-minimisation analysis' OR 'cost-minimization analysis' OR 'Cost consequence analysis' OR 'health care utility' OR 'Quality adjusted life year' OR 'Incremental Cost Effectiveness Ratio' OR 'quality-adjusted life year' OR 'quality-adjusted life-year' OR 'Cost effective' OR 'Cost-effective' OR 'Cost effectiveness' OR 'Cost-effectiveness' OR 'economic evaluation' OR 'economic*' OR 'Budget-impact model' OR 'Budget impact' OR 'Incremental-Cost-Effectiveness Ratio' OR 'Incremental-Cost Effectiveness Ratio' OR 'life years gained' OR 'life-years gained' OR 'patient level' OR 'discrete event' OR 'LYG' OR 'economic models' OR 'Base case' OR 'incremental-cost' OR 'Willingness-to-pay' OR 'Time trade off' OR 'Time trade-off' OR 'Time-trade off' OR 'Decision theory' OR 'pharmacoeconomic*'):ti,ab	475,774
#8	Resource use/cost	('cost benefit analysis' OR 'cost' OR 'cost of illness' OR 'resource usage' OR 'resource consumption' OR 'resource' OR 'resource allocation' OR 'resource management' OR 'cost-missing' OR 'cost missing' OR 'Disease Burden' OR 'Burden of illness' OR 'Burden of sickness' OR 'Sickness Burden' OR 'burden of disease' OR 'length of stay' OR 'duration of stay' OR 'extended stay' OR 'prolonged stay' OR 'duration of stay'	1,191,110

		OR ‘duration of hospitalisation’ OR ‘duration of hospitalization’ OR ‘bed-days’ OR ‘bed days’ OR ‘re-admi*’ OR ‘readmi*’ OR ‘hospitali*’ OR ‘readmission’ OR ‘hospital readmission’ OR ‘ICU stay’ OR ‘ICU day’ OR ‘Specialist visit’ OR ‘Outpatient visit’ OR ‘GP visit’ OR ‘General practitioner visit’ OR ‘Emergency room visit’ OR ‘ER visit’ OR ‘cost of illness’ OR ‘hospital stay’):ti,ab	
#9	QoL/HSUV terms	(‘health status’ OR ‘statistical model’):ti,ab	75,652
#10	QoL/HSUV terms	(‘SF-36’ OR ‘SF36’ OR ‘SF 36’ OR ‘short form 36’ OR ‘short form-36’ OR ‘short-form 36’ OR ‘SF-12’ OR ‘short form-12’ OR ‘short form 12’ OR ‘SF12’ OR ‘short-form 12’ OR ‘SF 12’):ti,ab	46,308
#11	All QoL/HSUV terms	#9 OR #10	117,087
#12	Clinical (Dis population + Intervention)	#5 AND #6	8,429
#13	Clinical - 10 years	#5 AND #6 AND [2008-2019]/py	6,949
#14	Disease term + Interv + CEA - 10 years	#1 AND #6 AND #7 AND [2008-2019]/py	225
#15	Dis population + Resource - 10 years	#5 AND #8 AND [2008-2019]/py	2,429
#16	Dis population + QoL/HSUV - 10 years	#5 AND #9 AND [2008-2019]/py	68
#17	Total	#13 OR #14 OR #15 OR #16 AND [2008-2019]/py	9,060

NICE National Institute for Health and Care Excellence

Clinical expert statement

Treosulfan with fludarabine for malignant disease before allogeneic stem cell transplant [ID1508]

Thank you for agreeing to give us your views on this technology and its possible use in the NHS.

You can provide a unique perspective on the technology in the context of current clinical practice that is not typically available from the published literature.

To help you give your views, please use this questionnaire. You do not have to answer every question – they are prompts to guide you. The text boxes will expand as you type.

Information on completing this expert statement

- Please do not embed documents (such as a PDF) in a submission because this may lead to the information being mislaid or make the submission unreadable
- We are committed to meeting the requirements of copyright legislation. If you intend to include **journal articles** in your submission you must have copyright clearance for these articles. We can accept journal articles in NICE Docs.
- Your response should not be longer than 13 pages.

About you	
1. Your name	Dr Amit Patel
2. Name of organisation	Clatterbridge Cancer Centre NHS Foundation Trust
3. Job title or position	Consultant/Programme Director, Stem Cell Transplantation and Cellular Therapy

4. Are you (please tick all that apply):	☐ an employee or representative of a healthcare professional organisation that represents clinicians? ☒ a specialist in the treatment of people with this condition? ☐ a specialist in the clinical evidence base for this condition or technology? ☐ other (please specify):
5. Do you wish to agree with your nominating organisation's submission? (We would encourage you to complete this form even if you agree with your nominating organisation's submission)	☒ yes, I agree with it ☐ no, I disagree with it ☐ I agree with some of it, but disagree with some of it ☐ other (they didn't submit one, I don't know if they submitted one etc.)
6. If you wrote the organisation submission and/ or do not have anything to add, tick here. (If you tick this box, the rest of this form will be deleted after submission.)	☐
The aim of treatment for this condition	

7. What is the main aim of treatment? (For example, to stop progression, to improve mobility, to cure the condition, or prevent progression or disability.)	Children and adults with blood cancers that are not deemed to be curable with chemotherapy and or radiotherapy alone, or have a high relapse risk, may be offered cellular immunotherapy in the form of allogeneic stem cell transplantation. The allogeneic donor transplant platform offers a chance of long term remission from the underlying cancer. This is thought to be achieved through a donor graft response raised against the host's blood cancer (graft versus tumour effect) which is thought to maintain (or achieve) long term disease control and lead to cure. Details of process are described further in the patient expert perspectives. Stem cell transplantation has a number of steps. Conditioning is one key element. It comprises a combination of immune suppressive and cytotoxic chemotherapy with or without radiotherapy. The aim of this is to remove recipient haematopoietic cells from the bone marrow (and residual cancer cells if still present), and prevent donor graft failure or rejection. Thus conditioning prevents the host from rejecting donor stem cells and clears the bone marrow to permit stable engraftment. This process is disease agnostic. Although most blood cancers have and are treated with this type of transplant, the most common disease indications for transplant in the UK are acute myeloid leukaemia (AML), myelodysplastic syndrome (MDS), and acute lymphoblastic leukaemia (ALL). Conditioning regimens can be divided in to myeloablative (MA) and non-myeloablative (NMA).[7.1]. Myeloablative regimens are expected to produce profound pancytopenia and myeloablation that is long lasting, usually irreversible, and in most instances fatal, unless haematopoiesis is restored by donor haemopoietic stem cell infusion. In contrast, a non-myeloablative regimen is expected to cause minimal cytopenia and does not necessary require stem cell support, but donor cell infusion reduces the period of cytopenia. Reduced intensity conditioning regimens (RIC) fall between these two definitions. RIC regimens differ from NMA as they cause cytopenia, which may be prolonged, and do require donor stem cell support. It is possible that autologous recovery would eventually occur after NMA conditioning, although pancytopenia would be of such duration to cause significant morbidity and mortality. Autologous recovery is typically not expected in MA regimens. Reduced toxicity regimens can be myeloablative and treosulfan regimens are an example. Thus, the choice of alkylator used within conditioning regimens is a key consideration.

	The main aim of a conditioning regimen is to balance transplant success (lack of graft failure), disease control (lack of relapse or progression), and toxicity (transplant related or non-relapse mortality, particularly from graft versus host disease and infections). Event-free survival (EFS) is thus a good surrogate endpoint to assess the efficacy and safety of conditioning regimens. Thus, taking the main aim of treatment in to account, an event is defined as graft failure, relapse, or death (whichever comes first). It typically takes 12-24 months to recover from a transplant. Relapses after 2 years are infrequent and would be unusual after 5 years, depending on the disease being treated. Patients that develop post transplant complications such as graft versus host disease (GvHD), infections, or both (typically co-exit) suffer significant morbidity, reduced quality of life, frequent hospital admissions/attendances, and mortality. The ideal conditioning regimen would thus be one that delivers MA but reduced toxicity conditioning with the highest EFS and overall survival, with the lowest rate of severe acute or chronic GvHD and infections. Use of the alkylator treosulfan instead of others such as busulfan or melphalan seems to achieve this best. Results from the phase 3 randomised controlled trial, one of the largest available in the EU and elsewhere, confirms patient benefit of treosulfan in a MAC regimen. Importantly, treosulfan outperformed a RIC dose busulfan regimen that is commonly used in the UK. [7.1] Bacigalupo A et al. Defining the Intensity of Conditioning Regimens: Working Definitions. *Biol Blood Marrow Transplant.* 2009 Dec;15(12):1628-33. doi: 10.1016/j.bbmt.2009.07.004.
8. What do you consider a clinically significant treatment response? (For example, a reduction in tumour size by x cm, or a reduction in disease activity by a certain amount.)	Improvement in event-free survival (EFS) is a good clinically significant endpoint to assess the efficacy and safety of treosulfan within a MAC regimen. Overall survival benefit at 24 months is also important, particularly if achieved with a lower rate of severe acute or chronic GvHD and infections; these events are associated with reduced patient experience as described by the submitted patient perspectives. The data reported in [8.1] represent a clinically significant treatment response, in a large randomised trial, using an alkylator used in the UK, at our centre and others. [8.1] Beelen DW et al. Treosulfan or busulfan plus fludarabine as conditioning treatment before allogeneic haemopoietic stem cell transplantation for older patients with acute myeloid leukaemia or myelodysplastic syndrome (MC-FludT.14/L): a randomised, non-inferiority, phase 3 trial. Lancet Haematol. 2019 Oct 9. pii: S2352-3026(19)30157-7. doi: 10.1016/S2352-3026(19)30157-7.

9. In your view, is there an unmet need for patients and healthcare professionals in this condition?	There is no universally agreed conditioning regimen for patients that require allogeneic stem cell transplantation. A desirable conditioning regimen is one that delivers the best efficacy and safety, measured by EFS and overall survival, and associated with improved patient experience related to reduced severe GvHD and infections. There is an unmet need from a reduced toxicity MAC regimen, and treosulfan seems to be the ideal alkylator for this purpose. Deaths from toxicity represent patients that may have potentially died unnecessarily from the transplant, that may not have relapsed, and that did not benefit from the morbidity associated with conditioning and subsequent recovery. Reducing toxicity associated with conditioning is a key unmet patient and healthcare professional need to improve patient experience and outcomes in the UK to match those in other parts of the world.
What is the expected place of the technology in current practice?	
10. How is the condition currently treated in the NHS?	Blood cancers treated with allogeneic stem cell transplantation would require conditioning chemotherapy and or radiotherapy. It is desirable to deliver MAC but excess toxicity in older and or co-morbid patients limits this approach due to excess mortality from the transplant itself. Thus, some patients receive RIC with the aim of reducing toxicity and thus improving overall survival. Treosulfan based conditioning in Beelen et al, and in other treosulfan studies outlined in the Medac submission, is myeloablative reduced toxicity conditioning. This extends the possibility of safely offering MAC to older and or co-morbid patients, improving EFS and overall survival while at the same time reducing complications like severe GvHD that are known to be associated with reduced patient experience and high healthcare utilisation [10.1] https://www.england.nhs.uk/wp-content/uploads/2017/03/gvhd-heamatopoietic-stem-cell.pdf In the UK, busulfan or melphalan might be used in RIC regimens. At our centre, we compared acute GvHD and other transplant and non-relapse mortality outcomes with MAC treosulfan, in a mixed population of patients with blood cancers. Similar to the Beelen et al trial and other registry data from EBMT and CIBMTR, treosulfan MAC had improved GvHD and reduced infection (viral reactivation) compared to melphalan based RIC [10.2]. These improved outcomes with treosulfan are consistent with other studies presented in the Medac submission, with the same signal of reduced toxicity and thus avoidable mortality.

	[10.1] Joseph Pidala J et al. Patient-reported quality of life is associated with severity of chronic graft-versus-host disease as measured by NIH criteria: report on baseline data from the Chronic GVHD Consortium. Blood. 2011 Apr 28; 117(17): 4651–4657. doi: 10.1182/blood-2010-11-319509 [10.2] Patel A et al. Elimination of Acute GvHD with Three Doses of Abatacept Prophylaxis Combined with Post-Transplant Cyclophosphamide after Myeloablative Treosulfan-Based Conditioning in the HLA Identical Sibling and Unrelated (10-12/12) Peripheral Blood Stem Cell Donor Setting: Improved Safety and Efficacy Compared to Alemtuzumab Prophylaxis. Data will be presented at the American Society of Hematology (ASH) Annual Meeting in December 2019. https://ash.confex.com/ash/2019/webprogram/Paper130184.html
• Are any clinical guidelines used in the treatment of the condition, and if so, which?	There are no universally accepted guidelines on allogeneic transplant conditioning. Each UK centre develops a conditioning regimen standard operating procedure (SOP) deliverable to their patient population within their facilities/resources, in compliance with internally agreed JACIE consensus standards. These SOPs are based on available evidence; Beelen et al would present a significantly large trial with sufficient evidence to adopt treosulfan based conditioning regimens in UK centres for blood cancer indications.
• Is the pathway of care well defined? Does it vary or are there differences of opinion between professionals across the NHS? (Please state if your experience is from outside England.)	The patient pathway of care is well defined in the NHS. Transplant centres a specially commissioned by NHS England and comply international quality standards required by JACIE accredited allogeneic stem cell transplant centres: http://bsbmt.org/uk-transplant-centre-list/ https://www.england.nhs.uk/wp-content/uploads/2018/07/Haematopoietic-stem-cell-transplantation-All-Ages.pdf Paediatric and adult disease indications for allogeneic stem cell transplant are also outlined by British Society of Blood and Marrow Transplantation (BSBMT) and NHS England: http://bsbmt.org/indications-table/ Patients are referred for consideration of allogeneic stem cell transplant by a referring consultant, typically after an MDT discussion. They are assessed by a transplant consultant and team, then after disease,

	donor, organ and fitness/patient performance status assessments are considered (among other factors), a conditioning regimen is chosen, and a transplant is planned. For younger fit patients MAC condition may be chosen, particularly if there is a higher risk of disease relapse. MAC is typically chosen if feasible in children and young adult patients. For older or co-morbid or less fit patients, or those with a lower disease relapse risk, a RIC regimen may be chosen. A RIC may be preferable for some disease groups, for example, lymphoma. Decision making is complex and transplant physicians and centres balance multiple factors to make decisions on the type of conditioning and the alkylator to use within the regimen. Treosulfan will permit delivery of reduced toxicity MAC to more patients what would otherwise only be suitable for RIC and improve the outcomes of patients receiving other MAC regimens associated with higher toxicity and consequent mortality.
• What impact would the technology have on the current pathway of care?	The current pathway of care described above is unlikely to change. Transplant conditioning approaches are likely to evolve to provide patients with access to treosulfan. This would be expected to improve patient outcomes such as survival and reduce toxicity. It would likely result in improved transplant efficiency and cost savings. These models are presented in the Medac submission.
11. Will the technology be used (or is it already used) in the same way as current care in NHS clinical practice?	Treosulfan is already available to use in MAC regimens in the UK, at our centre and others, and funded through the current NHS specialist commissioning tariff; this funds the first 100 days of the transplant and includes the conditioning regimen costs. Given that treosulfan does not represent an increase in acquisition costs, as awareness of the recently published Beelen et al data increases, more centres are likely to use this safer alkylator. Furthermore, as patient outcomes and experience are the major driver, and these are comparatively benchmarked for each centre in the UK, and soon across the Europe, it is likely that treosulfan use in the UK will increase to compete with the with improved EFS and survival outcomes that have been achieved in European centres with treosulfan (EBMT and other data in Medac submission).
• How does healthcare resource use differ between the technology and current care?	Treosulfan is associated with reduced toxicity, including from the major non-relapse side effects of transplant which are graft versus host disease and infection. Avoiding these post transplant events, and reducing their severity, would be associated with reduce resource utilisation and cost.

• In what clinical setting should the technology be used? (For example, primary or secondary care, specialist clinics.)	In NHS England commissioned, JACIE accredited, allogeneic stem cell transplant centres. http://bsbmt.org/uk-transplant-centre-list/ https://www.ebmt.org/jacie-accredited-centres
• What investment is needed to introduce the technology? (For example, for facilities, equipment, or training.)	No new investment was necessary to introduce treosulfan based conditioning at our centre. This would also be the same expectation across the NHS given that conditioning chemotherapy is included in the specialist commissioned transplant tariff. All transplant centres will be able to provide training relating to changes in their standard operating procedures (SOPs), and the conditioning regimen would present such a change. This is a key component to maintain JACIE accreditation and NHS England commissioning.
12. Do you expect the technology to provide clinically meaningful benefits compared with current care?	Yes, these are described above and demonstrated in Beelen et al and our UK centre abstract above. These are also described in detail within the Medac submission: improved survival, EFS, reduced serve GvHD and infections, and consequently improved patient experience.
• Do you expect the technology to increase length of life more than current care?	Yes. Overall survival was higher in the treosulfan treated patients compared to busulfan treated patients in Beelen et al: 72.7% (95% CI 66.8-77.8) compared to 60.2% (95% CI 54-65.8), P<0.05.
• Do you expect the technology to increase health-related quality of life more than current care?	Yes. The non-relapse events that were lower in the treosulfan group (graft failure, severe GvHD, infections) are all associated with reduced patient experience and quality of life. One would expect adoption of treosulfan into routine clinical practice to improve these patient outcomes from a fall in the undesirable post transplant non-relapse events.

13. Are there any groups of people for whom the technology would be more or less effective (or appropriate) than the general population?	Treosulfan would be a better choice than busulfan for patients with liver impairment and seizure disorders. This would also be the case for patient requiring treatment or exposure to azole class anti-fungal agents, or therapy/prophylaxis against mycobacterial infections which are often hepatotoxic. Patients requiring high doses of other interacting potentially hepatotoxic medications, including anti-depressants or strong analgesics, might also benefit from avoiding busulfan. Treosulfan would be the alkylator of choice for patients are risk of gut toxicity or mucositis, expected to be higher if melphalan is used. The combination of melphalan with other agents such as methotrexate would be expected to exacerbate these gut (and liver) directed complications. Treosulfan would be preferable to other alkylators if patients are at increased risk of veno-occlusive disease (VOD), a recognised post transplant adverse event associated with morbidity, increase in treatment and health utilisation costs, and mortality. Examples of patients are risk VOD include those with acute myeloid leukaemia (AML) or acute lymphoblastic leukaemia (ALL) treated with Gemtuzumab or inotuzumab respectively. Typically these patients may have relapsed disease or disease at high risk of relapse. Treosulfan use at our centre has been associated with a reduction in VOD rates requiring defibrotide use, according to NHS England guidance. https://www.england.nhs.uk/wp-content/uploads/2018/07/Defibrotide-in-severe-veno-occlusive-disease.pdf
The use of the technology	
14. Will the technology be easier or more difficult to use for patients or healthcare professionals than current care? Are there any practical implications for its use (for example, any concomitant	Treosulfan is easier to use compared to busulfan, particularly if dose adjustment regimens are used, and melphalan. Treosulfan is associated with reduced liver and neurological toxicity compared to thiotepa. No anti-epileptic therapy is required with treosulfan like it is for busulfan, there is a lower rate of organ impairment, there is a lower rate of severe mucositis in our experience, with GvHD and secondary infection from treatment. Treosulfan concentrates highly in the bone marrow and spleen, both desirable for transplant conditioning.

treatments needed, additional clinical requirements, factors affecting patient acceptability or ease of use or additional tests or monitoring needed.)	Treosulfan administration is also convenient as a peripheral cannula can be used if central access is not possible.
15. Will any rules (informal or formal) be used to start or stop treatment with the technology? Do these include any additional testing?	Not applicable as treosulfan is part of an allogeneic stem cell transplant conditioning regimen.
16. Do you consider that the use of the technology will result in any substantial health-related benefits that are unlikely to be included in the quality-adjusted life year (QALY) calculation?	Yes. Improved patient tolerance, reduced readmission to hospital with post transplant events including GvHD and improved patient satisfaction. The patient expert submissions also consider the psychological negative effects of these post transplant events , that occur some months after the procedure. Treosulfan is associated with reduced complication events and reduced severity (GvHD and infections). The beneficial psychological impact of treosulfan in relation to these may not be included in the QALY calculations.

17. Do you consider the technology to be innovative in its potential to make a significant and substantial impact on health-related benefits and how might it improve the way that current need is met?	Yes. Improved patient survival would be expected post allogeneic stem cell transplant for blood cancers in children and adults. Post transplant recovery would be expedited, with higher and more rapid attainment of chimerism, reduced incidence and severity of GvHD and associated infections.
• Is the technology a 'step-change' in the management of the condition?	Yes. It is rare that a large RCT has demonstrated a clear survival benefit of an alkylator as part of a conditioning regimen compared to another agent within a regimen. Recruitment to conditioning trials has always been challenging thus the Beelen et al study represents a significant robust body of evidence supporting treosulfan use. The magnitude of survival and EFS benefit represents a step change improvement in patient outcomes that are clinically important.
• Does the use of the technology address any particular unmet need of the patient population?	Yes. Patient outcomes post allogeneic stem cell transplant requirement improvement in UK centres, with a reduction in mortality.
18. How do any side effects or adverse effects of the technology affect the	Compared to other alkylators such as busulfan or melphalan, treosulfan based regimens have reduced side effects such a s GvHD and infections. Patient experience and quality of life would be expected to be better

management of the condition and the patient's quality of life?	with treosulfan regimens, and this is our centre's experience compared to melphalan, busulfan, and other MAC and RIC regimens.
Sources of evidence	
19. Do the clinical trials on the technology reflect current UK clinical practice?	Yes, as described above.
• If not, how could the results be extrapolated to the UK setting?	N/A
• What, in your view, are the most important outcomes, and were they measured in the trials?	The key outcomes were measured and reported; EFS, survival, engraftment/graft failure, chimerism, TRM, NRM, relapse, severe acute and chronic GvHD, infections.
• If surrogate outcome measures were used, do they adequately predict long-term clinical outcomes?	Yes, EFS and survival, as described above.
• Are there any adverse effects that were not apparent in clinical trials	No.

but have come to light subsequently?	
20. Are you aware of any relevant evidence that might not be found by a systematic review of the trial evidence?	No.
21. Are you aware of any new evidence for the comparator treatment(s) since the publication of NICE technology appraisal guidance	No.
22. How do data on real-world experience compare with the trial data?	The Beelen et al trail data is consistent with EBMT and CIBMTR registry data and investigator initiated/sponsored studies. These other studies are outlined in the Medac submission. Our UK data are also consistent, compared against a melphalan RIC regimen.[22.1] [22.1] Patel A et al. Elimination of Acute GvHD with Three Doses of Abatacept Prophylaxis Combined with Post-Transplant Cyclophosphamide after Myeloablative Treosulfan-Based Conditioning in the HLA Identical Sibling and Unrelated (10-12/12) Peripheral Blood Stem Cell Donor Setting: Improved Safety and Efficacy Compared to Alemtuzumab Prophylaxis. Data will be presented at the American Society of Hematology

	(ASH) Annual Meeting in December 2019. https://ash.confex.com/ash/2019/webprogram/Paper130184.html
Equality	
23a. Are there any potential equality issues that should be taken into account when considering this treatment?	N/A.
23b. Consider whether these issues are different from issues with current care and why.	N/A
Key messages	

24. In up to 5 bullet points, please summarise the key messages of your statement.

- Improved patient survival outcomes post allogeneic stem cell transplant with treosulfan based myeloablative conditioning for allogeneic stem cell transplantation.
- Reduced toxicity with treosulfan despite MAC permitting delivery and patient benefit to older and co-morbid/less fit patients, as well as children and younger patients.
- Improved safety and tolerability, including in patients that might have pre-treatment organ impairment, for example liver impairment or neurological conditions such as epilepsy, reducing the risk of VOD.
- Improved patient experience and outcomes based on our UK cohort of adults, including elderly patients.
- Reduced need for post transplant readmissions to hospital and daycare due to reduced events/complications, including graft versus host disease and infections.

Thank you for your time.

Please log in to your NICE Docs account to upload your completed statement, declaration of interest form and consent form.

.....

Your privacy

The information that you provide on this form will be used to contact you about the topic above.

☐ **Please tick this box** if you would like to receive information about other NICE topics.

For more information about how we process your personal data please see our privacy notice.....

NICE National Institute for Health and Care Excellence

Patient expert statement

Treosulfan with fludarabine for malignant disease before allogeneic stem cell transplant [ID1508]

Thank you for agreeing to give us your views on this technology and its possible use in the NHS.

You can provide a unique perspective on conditions and their treatment that is not typically available from other sources.

To help you give your views, please use this questionnaire with our guide for patient submissions.

You do not have to answer every question – they are prompts to guide you. The text boxes will expand as you type.

Information on completing this expert statement

- Please do not embed documents (such as a PDF) in a submission because this may lead to the information being mislaid or make the submission unreadable
- We are committed to meeting the requirements of copyright legislation. If you intend to include **journal articles** in your submission you must have copyright clearance for these articles. We can accept journal articles in NICE Docs.
- Your response should not be longer than 10 pages.

About you	
1.Your name	**Caitlin Farrow**
2. Are you (please tick all that apply):	☐ a patient with the condition? ☐ a carer of a patient with the condition? ☒ a patient organisation employee or volunteer? ☐ other (please specify):

3. Name of your nominating organisation	**Anthony Nolan**
4. Did your nominating organisation submit a submission?	☐ yes, they did ☒ no, they didn't ☐ I don't know
5. Do you wish to agree with your nominating organisation's submission? (We would encourage you to complete this form even if you agree with your nominating organisation's submission)	☐ yes, I agree with it ☐ no, I disagree with it ☐ I agree with some of it, but disagree with some of it ☒ other (they didn't submit one, I don't know if they submitted one etc.)

6. If you wrote the organisation submission and/ or do not have anything to add, tick here. (If you tick this box, the rest of this form will be deleted after submission.)	☐ yes
7. How did you gather the information included in your statement? (please tick all that apply)	☒ I am drawing on others' experiences. Please specify how this information was gathered: • Anthony Nolan created a survey for patients with malignant disease who had undergone conditioning prior to an allogeneic haematopoietic stem cell transplant. The survey focused on the conditioning regimen and associated side effects. This was shared with our Patients and Families Panel. • At the time of submission, there were 20 responses to this consultation; primarily from patients located in England, however there was a minority of patients from Wales and Scotland. • The survey provides a snapshot of the patient experience of pre-transplant conditioning. Given the small sample size, and that respondents were self-selecting, this is not representative of all experiences of alloHSCT. The submissions have been provided by a subset of our Patients and Families panel who volunteered to provide details of their personal experience, which in some cases may have been a number of years ago. • We also had informal discussion about the use of conditioning regimens with medical professionals to support this submission. **Our role representing patients** Anthony Nolan saves and improves the lives of people with blood cancers and blood disorders in need of a potentially curative stem cell transplant. We provide patients with matching donors from our stem cell donor register and facilitate their transplants, support them and their families throughout their transplant journey, and advocate on their behalf. We have over 760,000 potential donors on the Anthony Nolan Stem Cell Register. Our aim is to improve outcomes and quality of life for our patients and therefore, we believe it is important that patient experience of conditioning regimens and transplant are represented to NICE as part of this appraisal. However, the

	nature of this technology has presented challenges over how to best represent patients on this issue. These challenges stem from the fact that conditioning regimens are tailored to the specific patient, considering factors such as diagnosis and fitness. As such, patients' experiences will only speak to that regimen. Similarly, patients are unable to compare their own conditioning regimen with one they have not experienced. Furthermore, given that conditioning comes directly prior to the infusion of donor cells, some patients may struggle to separate the effects of a conditioning regimen from that of allogeneic transplant. Taking into account these challenges, we have written our submission to provide background knowledge of alloHSCT as an intervention, the role of conditioning and the way it is experienced by patients, rather than presenting patient opinion on the specific technology in question.
Living with the condition	
8. What is it like to live with the condition? What do carers experience when caring for someone with the condition?	**<u>Allogeneic Haematopoietic stem cell transplants and conditioning regimens</u>** Allogeneic haematopoietic stem cell transplantation (alloHSCT) is an intervention used to treat a wide spectrum of haematological disorders, including malignant disease. The most common clinical indications for HSCT are leukaemias, lymphomas and myeloma. AlloHSCT is a potentially curative therapy, meaning that the patient's blood and immune system becomes healthy again, curing the diagnosed blood cancer, and is able to fight infection. However, alloHSCT has many risks for the patient and is only considered as a treatment option for patients who would not benefit from prolonged disease management or whose disease is unlikely to be cured with chemotherapy alone. Before a patient undergoes alloHSCT, their immune system will be treated and prepared to make way for new cells, this is called conditioning. The role of conditioning, making room for the donor stem cells, necessarily incurs side effects in its use. There are different intensities of conditioning, classified as myeloablative, reduced-intensity, and nonmyeloablative. The intensity of conditioning regimens can vary substantially, as the optimal conditioning regimen for a patient will be determined by disease-related factors such as diagnosis and remission status, as well as patient-related factors including age, donor availability, and presence of comorbid conditions. This was borne out in the survey as, amongst the respondents who were able to recall their specific treatment, there was a variety of different regimens. All respondents were able to tell us if they had a reduced or full intensity regimen, however, with 11 of the 20 (55%) respondents treated with a reduced intensity regimen and 9 of the 20 (45%) with full intensity. Besides the risk of relapse of the disease, haematopoietic stem cell transplantation remains associated with significant early and late treatment related mortality (TRM). Infections, toxicity, and graft-vs.-host disease (GVHD) are

	the main causes of death.[1] An indication of transplant 'success' is whether the patient is alive at five years post-transplant. For allogeneic transplants, the five-year overall survival is 48% for adults and 73% for children.
Current treatment of the condition in the NHS	
9. What do patients or carers think of current treatments and care available on the NHS?	**The role of conditioning in transplant** • Patients told us that they understood the reason for undergoing conditioning treatment prior to transplant, and as such accepted it as a necessary part of their transplant journey. ○ *"I understood that the pre conditioning treatment had to be done as it was to give the best chance of a successful transplant*." **Side effects from conditioning regimens** • While patients will have experienced side effects differently dependent on their fitness prior to conditioning and the intensity of the regimen, all the respondents mentioned at least one side effect in their submission. • In over half of the responses, 13 out of 20 (65%), at least three different side effects were reported. • The most commonly reported were; extreme fatigue, nausea and vomiting, diarrhoea, loss of appetite, constipation, mucositis, ulcers and hair loss. • Side effects were often experienced in a very severe form, necessitating additional treatment, and in certain cases led to more extreme treatment-related morbidity. • For example, one patient reported that they experienced severe constipation which led to an acute haemorrhoid and anal fissure which caused them significant pain. The patient stated that they then had E. coli septicaemia, which had to be treated with intravenous antibiotics, which they believed was probably caused by the initial effects of the constipation. This case indicates the potential for side effects to escalate in this immunocomprised patient population. **Impact of side effects** • Despite the understanding of the role, and need, for this treatment, the actual experience of conditioning was a shock for many patients.

[1] https://www.nature.com/articles/s41409-019-0624-z, Death after hematopoietic stem cell transplantation: changes over calendar year time, infections and associated factors, EBMT 2019

	○ *"I found this part of the treatment pretty tough. I knew the reason for it, so I prepared mentally but the physical shock is quite hard."* ○ *"I realise it is a means to an end and fortunately transplant is immediately afterwards"* • The conditioning regimen, for many patients, was more challenging than their previous experiences of chemotherapy. ○ One response highlighted that the "*very strong nausea and vomiting*" they experienced from their reduced intensity conditioning regimen was significantly worse than their prior experience of chemotherapy. • For many patients, these side effects had a significant impact on their day to day life, with many describing them as very challenging: ○ *"Whilst undergoing the treatment, the side effects were horrendous. The side effects continued for quite a considerable time".* ○ *"The second course made me very ill and I was unable to eat properly for 6 weeks"* ○ "V*ery, very vicious treatment*". **Impact on quality of life** • Patients told us that the side effects from pre-transplant conditioning had a significant effect on their quality of life, their independence, including their ability to look after themselves, and their emotional health and wellbeing. • 19 of 20 (95%) patients surveyed said that the side effects of pre-transplant conditioning had a 'negative' or 'very negative' effect on their mental health and well-being. ○ *"I was completely bed ridden, with no quality of life experiencing many health issues, totally dependent of nursing staff, and well-being been supported by family*." ○ *"I lost confidence, feared being away from a toilet for too long, was room bound too for a lot of time."* ○ *"It was a worrying time and a struggle emotionally"* • Patients also reported that the experience of these side effects were challenging for their family and carers too. ○ *"My family found it difficult to cope and support me during this period.* *It caused a lot of stress because they didn't understand or know when things would improve*." • Patients also reported that the negative impact on their emotional health and well-being had, in many cases, outlasted the physical effects of their conditioning regimen.

	○ *I do still have vivid memories of the conditioning treatment, I think they will remain with me always.* ○ *This was a tough time and experience, leaving me to this day with many psychological effects*.
10. Is there an unmet need for patients with this condition?	**Future conditioning regimens** • The respondents told us that they understood the role of conditioning as necessary to their transplant and while the side effects experienced were not insignificant, they did not regret undergoing their transplant. ○ *"I was glad to be alive"* • However, patients told us that the introduction of any conditioning regimen with reduced toxicity and side-effect profile would be welcomed to benefit future patients. ○ *"I really hope new treatments can reduce the suffering of future patients- I was not prepared for the agony I was to endure to obliterate my faulty bone marrow."*
Advantages of the technology	
11. What do patients or carers think are the advantages of the technology?	NA – We did not have any responses from patients who had received the technology.
Disadvantages of the technology	
12. What do patients or carers think are the disadvantages of the technology?	NA – We did not have any responses from patients who had received the technology.
Patient population	
13. Are there any groups of patients who might benefit	NA – We have not explored the specific potential of this technology in our submission, but any conditioning regimen that is better tolerated by patients has the potential to enable patients to receive AlloHSCT where they might not have been able to previously.

more or less from the technology than others? If so, please describe them and explain why.	
Equality	
14. Are there any potential equality issues that should be taken into account when considering this condition and the technology?	NA – We have not identified any equality issues.
Other issues	
15. Are there any other issues that you would like the committee to consider?	No.
Key messages	
16. In up to 5 bullet points, please summarise the key messages of your statement: • Patients understand the role of conditioning as necessary to their transplant and while the side effects experienced were not insignificant, they did not regret undergoing their transplant	

- Patients experience side effects differently but all survey respondents mentioned at least one side effect, and up to seven, in their submissions.
- The side effects from pre-transplant conditioning had a significant effect on patient's quality of life, independence, including their ability to look after themselves, and their emotional health and wellbeing.
- Patients also reported that the negative impact on their emotional health and well-being had, in many cases, outlasted the physical effects of their conditioning regimen.
- Patients welcome the introduction of any conditioning regimen with reduced toxicity and side-effect profile which could benefit future patients.

Thank you for your time.

Please log in to your NICE Docs account to upload your completed statement, declaration of interest form and consent form.

NICE National Institute for Health and Care Excellence

Patient expert statement

Treosulfan with fludarabine for malignant disease before allogeneic stem cell transplant [ID1508]

Thank you for agreeing to give us your views on this technology and its possible use in the NHS.

You can provide a unique perspective on conditions and their treatment that is not typically available from other sources.

To help you give your views, please use this questionnaire with our guide for patient submissions.

You do not have to answer every question – they are prompts to guide you. The text boxes will expand as you type.

Information on completing this expert statement

- Please do not embed documents (such as a PDF) in a submission because this may lead to the information being mislaid or make the submission unreadable
- We are committed to meeting the requirements of copyright legislation. If you intend to include **journal articles** in your submission you must have copyright clearance for these articles. We can accept journal articles in NICE Docs.
- Your response should not be longer than 10 pages.

About you	
1.Your name	**Alan Tindale**

2. Are you (please tick all that apply):	A patient with the condition
3. Name of your nominating organisation	Anthony Nolan
4. Did your nominating organisation submit a submission?	☐ yes, they did
5. Do you wish to agree with your nominating organisation's submission? (We would encourage you to complete this form even if you agree with your nominating organisation's submission)	☐ yes, I agree with it

6. If you wrote the organisation submission and/ or do not have anything to add, tick here. (If you tick this box, the rest of this form will be deleted after submission.)	☐
7. How did you gather the information included in your statement? (please tick all that apply)	I have personal experience of the condition
Living with the condition	
8. What is it like to live with the condition? What do carers experience when caring for someone with the condition?	I was obviously very ill before diagnosis of Acute Lymphoblastic Leukaemia (ALL) and had several symptoms that presented themselves at different times. This illness affected my every day life in many ways as I was unable to function properly and I had no energy, drive or motivation. After many blood tests I was admitted to hospital and after many other tests I was diagnosed. In truth, it was a relief to receive confirmation of my worst fears as it meant that I/ the specialists could begin to fight the disease. After a prolonged period of hospitalisation I was recuperating at home waiting for the donor system to be activated. I had begun to feel better and I appeared to be making progress with my recovery. I was admitted to hospital again several days before my transplant was due. It was at this stage that I was given drugs to prepare me to receive the donation. Shortly after I received the first dose of drugs my body started to “rumble” from within my core and it culminated in a state of violent shakes that were scary and I find it difficult to accurately articulate how powerful they were. I felt poorly instantly and although I knew that I was being prepared for the actual transplant, I felt really low

	and irritated as it felt that I had taken a huge backward step in my recovery process. I was given the same drug the following day although the effects were not as severe on this occasion. After the transplant I was told to expect a stay in hospital of between 4 – 6 weeks until my levels were sufficiently high enough for me to be released. This time was a strange period of inactivity and just waiting for something to happen. I remember feeling relatively strong and it was apparent that I had reacted well to the transplant. My levels rose reasonably quickly and I was released from hospital after only eleven days. I was told that this was quite an extraordinary occurrence. Post transplant, I was relatively well although I did suffer from some Graft versus Host symptoms such as blisters in my mouth and rashes on my skin. This involved many trips back to hospital for check ups and to give blood samples etc. I did receive some after care at home but mostly to clean my HIC line and change any dressings if required. As time progressed these symptoms eased although there has been a definite weakening and vulnerability in my immune system. After a while I developed shingles in my head and was taken to hospital locally in an ambulance after nearly collapsing at home. I spent a week in hospital. I also developed pneumonia and spent three weeks in hospital, including two weeks in intensive care which was an extremely traumatic and worrying experience. I develop colds very easily and every winter since my transplant I spend several months appearing to pick up several bugs that leave me feeling down and very weak. Until last winter, I used to go to the Royal Marsden Hospital monthly to receive a three hour intravenous dose of the drug Immunoglobulin in an attempt to boost my immune system. Generally, I am a fit person and I lead a very active lifestyle. Until recently I still played rugby although now I only referee matches on a regular basis. I completed a marathon earlier this year on the 5th anniversary of my transplant.

Current treatment of the condition in the NHS	
9. What do patients or carers think of current treatments and care available on the NHS?	I was astounded at how good the treatment offered by the NHS was. I was treated exceptionally well whilst in hospital and I witnessed first hand the professionalism, care and dedication of all NHS staff that I came into contact with. I have nothing but admiration for them all. Being told you have a life changing/ threatening illness is naturally a devastating blow for (I would imagine) anybody but I took it particularly badly as I was an active soldier, sportsman and somebody who absolutely lived life to the full. The thought that I could have all of this taken away was devastating and I struggled to come to terms with the diagnosis initially. After a brief period of self pity, I determined myself to fighting to recover and, ultimately survival. I became ultra positive, a trait that seemed to be well received by the medical staff at the Royal Marsden Hospital. I was told that they had seen many patients who were either very pessimistic or who had simply "given up" in similar circumstances. I was very fortunate to have an amazingly dedicated and resilient life partner (my wife) by my side throughout and an immensely supportive welfare structure in place (the Army) offering me support along the way. During my time in hospital I developed friendships with the wonderful people who cared for me and I had plenty of time to actually appreciate how hard they work on a daily basis. My subsequent stays in my local hospital (Swindon) also demonstrated how stretched the nursing staff are and how much they have to try and achieve during any given shift. I witnessed the NHS "service" from the A & E entry phase through to general and intensive care. I was astounded at the difference in the level of care that I received in the different areas – this is not a criticism of any of the staff but an observation on the amount that they have to do and the stuff that they have to contend with.
10. Is there an unmet need for patients with this condition?	I believe that there was a severe lack of psychological support/ awareness around my treatment. Being informed that you have cancer is a life changing situation for both the patient and family members. It seemed that there was a lack of personnel equipped to deal with the psychological elements of dealing with a critical illness diagnosis. I found this to be the case throughout my treatment, my transplant and beyond. I was also disappointed at the lack of contact with MacMillan; having seen so many advertisements about their involvement with cancer care, I was not contacted once throughout my treatment or once I returned home.

Advantages of the technology	
11. What do patients or carers think are the advantages of the technology?	Any drug without side effects would be good. I found the whole preparation phase (taking drugs) traumatic and a major set back in the recovery process.
Disadvantages of the technology	
12. What do patients or carers think are the disadvantages of the technology?	
Patient population	
13. Are there any groups of patients who might benefit more or less from the technology than others? If so, please describe them and explain why.	
Equality	
14. Are there any potential equality issues that should be	

taken into account when considering this condition and the technology?	
Other issues	
15. Are there any other issues that you would like the committee to consider?	I feel that a better system in place to deal with the psychological effects that the *whole* process has on patients and families alike.
Key messages	
16. In up to 5 bullet points, please summarise the key messages of your statement: • Psychological support required throughout the process. • Preparatory (pre transplant) drugs not have such negative effects on the patient. • NHS staff and systems are incredible for specialist illness(es). • More information regarding the whole recovery process should be made available. • Specific after care support could be improved.	

Thank you for your time.

Please log in to your NICE Docs account to upload your completed statement, declaration of interest form and consent form.

.....

Your privacy

The information that you provide on this form will be used to contact you about the topic above.

☐ **Please tick this box** if you would like to receive information about other NICE topics.

For more information about how we process your personal data please see our privacy notice.

.....

in collaboration with:

Erasmus School of Health Policy & Management

Maastricht University

Treosulfan with fludarabine for malignant disease before allogeneic stem cell transplant

Produced by Kleijnen Systematic Reviews (KSR) Ltd. in collaboration with Erasmus University Rotterdam (EUR) and Maastricht University

Authors

Marie Westwood, Reviews Manager, KSR Ltd, UK
Isaac Corro Ramos, Health Economics Researcher, Institute for Medical Technology Assessment, EUR, The Netherlands
Robert Wolff, Deputy Director, KSR Ltd, UK
Thea Drachen, Systematic Reviewer, KSR Ltd, UK
Hannah Penton, Health Economics Researcher, Erasmus School of Health Policy and Management, EUR, The Netherlands
Pim Wetzelaer, Health Economics Researcher, Erasmus School of Health Policy and Management, EUR, The Netherlands
Nigel Armstrong, Health Economics Manager, KSR Ltd, UK
Titas Buksnys, Health Economist, KSR Ltd, UK
Gill Worthy, Statistician, KSR Ltd, UK
Shelley de Kock, Information Specialist, KSR Ltd, UK
Maiwenn Al, Health Economics Researcher, Erasmus School of Health Policy and Management, EUR, The Netherlands
Jos Kleijnen, Director, KSR Ltd, UK; Professor of Systematic Reviews in Health Care, Maastricht University, The Netherlands

Correspondence to Robert Wolff, Kleijnen Systematic Reviews
Unit 6, Escrick Business Park
Riccall Road, Escrick
York, UK
YO19 6FD

Date completed 29/07/2019

Source of funding: This report was commissioned by the NIHR HTA Programme as project number 12/91/81.

Declared competing interests of the authors
None.

Acknowledgements
None.

Commercial in confidence (CiC) data are highlighted in blue throughout the report.

Academic in confidence (AiC) data are highlighted in yellow throughout 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:
Westwood M, Corro Ramos I, Wolff R, Drachen T, Penton H, Wetzelaer P, Armstrong N, Buksnys T, Worthy G, de Kock S, Al M, Kleijnen J. Treosulfan with fludarabine for malignant disease before allogeneic stem cell transplant: a Single Technology Assessment. York: Kleijnen Systematic Reviews Ltd, 2019.

Contributions of authors
Marie Westwood and Robert Wolff acted as joint project leads and systematic reviewers on this assessment, critiqued the clinical effectiveness methods and evidence and contributed to the writing of the report. Isaac Corro Ramos acted as health economic project lead, critiqued the company's economic evaluation and contributed to the writing of the report. Hannah Penton, Pim Wetzelaer, Titas Buksnys and Nigel Armstrong acted as health economists on this assessment, critiqued the company's economic evaluation and contributed to the writing of the report. Thea Drachen acted as systematic reviewer, critiqued the clinical effectiveness methods and evidence and contributed to the writing of the report. Gill Worthy acted as statistician, critiqued the analyses in the company's submission and contributed to the writing of the report. Shelley de Kock critiqued the search methods in the submission and contributed to the writing of the report. Maiwenn Al acted as health economist on this assessment, critiqued the company's economic evaluation, contributed to the writing of the report and provided general guidance. Jos Kleijnen critiqued the company's definition of the decision problem and their description of the underlying health problem and current service provision, contributed to the writing of the report and supervised the project.

Abbreviations

AE	Adverse event
aGvHD	Acute graft-versus-host disease
AIC	Akaike information criterion
AiC	Academic in confidence
ALL	Acute lymphoblastic leukaemia
alloHSCT	Allogeneic haematopoietic stem cell transplantation
AML	Acute myeloid leukaemia
APL/APML	Acute promyelocytic leukaemia
ASBMT	American Society for Blood and Marrow Transplantation
ASH	American Society of Haematology
ATG	Anti-thymocyte globulin
BIC	Bayesian information criterion
BM	Bone marrow
BNF	British National Formulary
BSBMT	British Society of Blood and Marrow Transplantation
BSA	Body surface area
Bu	Busulfan
CE	Conformité Européenne (engl. European conformity)
CE	Cost effectiveness
CEAC	Cost effectiveness acceptability curve
cGvHD	Chronic graft-versus-host disease
CHMP	Committee for Medicinal Products for Human Use
CI	Confidence interval
CIBMTR	US Center for International Blood and Marrow Transplant Research
CiC	Commercial in confidence
CLL	Chronic lymphoblastic leukaemia
CML	Chronic myeloid leukaemia
CMML	Chronic myelomonocytic leukaemia
CNS	Central nervous system
CNS	Clinical nurse specialist
CR1	First complete remission
CR	Complete remission
CRF	Case report form
CRO	Contract research organisation
CRP	C-reactive protein
CS	Company submission
CSR	Clinical study report
CT	Chemotherapy
CTC	Common Terminology Criteria
CTCAE	Common Terminology Criteria for Adverse Events (National Cancer Institute)
Cy	Cyclophosphamide
DCE	Discrete choice experiment
DFS	Disease-free survival
DLCOSB	Single breath diffusing capacity for carbon monoxide
DLI	Donor lymphocyte infusion
DMC	Data monitoring committee
DNA	Deoxyribonucleic acid
DSU	Decision Support Unit
EBV	Epstain-Barr virus
ECG	Electrocardiogram
EFS	Event-free survival
EHA	European Hematology Association
EBMT	European Society for Blood and Marrow Transplantation

EPAR	European Public Assessment Report
EQ-5D	European Quality of Life-5 Dimensions
ERG	Evidence Review Group
EUR	Erasmus University Rotterdam
FAS	Full analysis set
FEV_1	Forced expiratory volume 1 second
Flag/IDA	Fludarabine/cytarabine/granulocyte-colony stimulating factor/idarubicin
Flu	Fludarabine
g	Gram
G-CSF	Granulocyte colony-stimulating factor
GGT	Gamma glutamyl transferase
GM-CSF	Granulocyte-macrophage colony-stimulating factor
GmbH	Gesellschaft mit beschränkter Haftung [engl. company with limited liability]
GvHD	Graft-versus-host disease
HCHS	Hospital and Community Health Services
HCRU	Health care resource utilisation
HCT-CI	Haematopoietic cell transplantation co-morbidity index
HERC	Health Economics Research Centre
HIV	Human immunodeficiency virus
HL	Hodgkin lymphoma
HLA	Human leukocyte antigen
HR	Hazard ratio
HRG	Healthcare resource group
HRQoL	Health-related quality of life
HSOS	Hepatic sinusoidal obstruction syndrome
HSCT	Haematopoietic stem cell transplantation
HSUV	Health state utility value
HTA	Health technology assessment
ICER	Incremental cost effectiveness ratio
Incr.	Incremental
INMB	Incremental net monetary benefit
IPSS-R	International Prognostic Scoring System-Revised
ITT	Intention to treat
i.v.	Intravenous
JMML	Juvenile myelomonocytic leukaemias
KM	Kaplan–Meier
KOL	Key opinion leader
KPS	Karnofsky performance status
KSR	Kleijnen Systematic Reviews
LVEF	Left ventricular ejection fraction
LYG	Life year gained
MAC	Myeloablative conditioning
MAE	Mean absolute error
MCM	Mixture-cure model
MDS	Myelodysplastic syndrome
MEC	Mitoxantrone, etoposide & cytarabine
Mel	Melphalan
mg	Milligram
MM	Multiple myeloma
MPAL	Mixed phenotype acute leukaemia
MPN	Myeloproliferative neoplasm
MRD	Matched related donor
MTD	Maximum tolerated dose
MUD	Matched unrelated donor
N	Number of participants

n	Number of events
N/A	Not applicable
NHL	Non-Hodgkin lymphoma
NHS	National Health Services
NICE	National Institute for Health and Care Excellence
NIHR	National Institute for Health Research
NK	Natural killer
NMA	Non-myeoablative
NMCM	Non-mixture-cure model
NOS	Not otherwise specified
NRM	Non-relapse mortality
OECD	Organisation for Economic Co-operation and Development
OS	Overall survival
PB	Peripheral blood
PCT	Procalcitonin
PDF	Portable document format
PFS	Progression-free survival
PPP	Purchasing power parity estimate
PPS	Per protocol set
PSA	Probabilistic sensitivity analyses
PSS	Personal Social Services
PSSRU	Personal Social Services Research Unit
QALY(s)	Quality-adjusted life year(s)
QLQ-C30	Quality of life core questionnaire C30
RBC	Red blood cells
RCT	Randomised controlled trial
RG	Risk group
rHU-KGF	Recombinant human keratinocyte growth factor
RI	Relapse incidence
RIC	Reduced intensity conditioning
RMSE	Root mean square error
RNA	Ribonucleic acid
RTC	Reduced toxicity conditioning
RR	Relapse rate
SAE	Serious adverse events
SAP	Statistical analysis plan
SAS	Safety analysis set
SD	Standard deviation
SE	Standard error
SmPC	Summary of product characteristics
SMR	Standardised mortality ratio
SpR	Specialist registrar
STA	Single technology appraisal
ULN	Upper limit of normal
TA	Technology appraisal
TBI	Total body irradiation
TEAE	Treatment-emergent adverse event
THIO	Thiotepa
TREO	Treosulfan
TRM	Transplantation-related mortality
TTO	Time trade-off
UK	United Kingdom
vs	Versus
WHO	World Health Organization
WTP	Willingness-to-pay

Table of Contents

Table of Tables

Table of Figures

1. Executive summary

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

Overall, the Evidence Review Group (ERG) agrees with the description of the underlying health problem. However, as detailed in Section 3 of this report, the ERG has a number of concerns about how this was implemented in the decision problem, most critically with regards to the population definition.

The population presented in the company submission (CS) is a subgroup of that defined in the scope and covered in the licenced indication for treosulfan. The submission relies, primarily, on one randomised controlled trial (RCT) of treosulfan (MC-FludT.14/L Trial II) which assessed a population that was narrower than that defined in the final scope:

- The final scope issued by the National Institute for Health and Care Excellence (NICE) defined the population of interest to be *"People with malignant disease that is in remission before allogenic haematopoietic stem cell transplantation"*.
- In line with the scope, the decision problem in the CS defined the population as *" Adults, children and young people with malignant disease that is in remission before allogenic haematopoietic stem cell transplantation".* The CS stated that *"this technology appraisal evaluates the clinical and cost-effectiveness of treosulfan as a conditioning treatment for malignant disease prior to allogeneic haematopoietic stem cell transplantation (alloHSCT) in paediatric and adult patients older than one month"*.
- However, MC-FludT.14/L Trial II only assessed the efficacy and safety of treosulfan in combination with fludarabine for conditioning therapy as part of alloHSCT in adults with acute myeloid leukaemia (AML) or myelodysplastic syndrome (MDS) who were at increased risk for standard conditioning therapies (i.e. not eligible for standard myeloablative conditioning (MAC) busulfan- or TBI-based regimens).

As presented in Section 3.1, the ERG asked for clarification on these issues. However, the company response was unsatisfactory in that it relies on a letter of support from clinical experts as the basis for assuming that the population addressed in the MC-FludT.14/L Trial II can be expanded to that defined in the final scope.

The ERG remains very concerned as there is a lack of evidence about the effectiveness of treosulfan/fludarabine conditioning regimens in people who are able to tolerate standard MAC regimens, in adults with malignancies other than AML or MDS, and in children.

Furthermore, the nature of the restricted population in MC-FludT.14/L Trial II (adults who were at increased risk and therefore not eligible for standard MAC conditioning regimens), and hence the choice of comparator (reduced intensity conditioning busulfan/fludarabine; RIC Bu/Flu) means that the relative effectiveness and cost effectiveness of treosulfan/fludarabine has not been evaluated against the range of comparator regimens that would be relevant for the full population defined in the scope.

1.2 Summary of the key issues in the clinical effectiveness evidence

The systematic review, used to identify studies for inclusion in the CS, was generally well conducted.

A range of resources were searched. Searches were well documented making them transparent and reproducible. Supplementary searches of conference proceedings were also undertaken to find unpublished and ongoing studies for the related topics. Appropriate subject headings and free text terms

were applied and combined correctly with Boolean logic for each facet in the search strategies. An RCT filter was not applied to PubMed and Cochrane Library searches and this is preferable, when possible, as it reduces the risk of missing relevant studies particularly those relating to adverse events.

Searches for interventions and comparators would have benefited from inclusion of more synonyms. For example, Embase searches did not include the term "ovastat" which is a brand name for treosulfan; Embase, PubMed and Cochrane Library searches did not include "oforta" which is a brand name for fludarabine or "evomela" which is a brand name for melphalan.

The evidence for treosulfan in combination with fludarabine for conditioning therapy as part of alloHSCT is based on a multi-centre, international RCT (MC-FludT.14/L Trial II) investigating patient-relevant outcomes, with follow-up to two years after transplantation (see Tables 1.1 and 1.2).

The population included in MC-FludT.14/L Trial II is not representative of the full population defined in the scope, specifically, there is a lack of evidence about the effectiveness of treosulfan/fludarabine (Treo/Flu) conditioning regimens in people who are able to tolerate standard MAC regimens, in adults with malignancies other than AML or MDS, and in children.

Furthermore, the nature of the restricted population in MC-FludT.14/L Trial II (adults who were at increased risk and therefore not eligible for standard MAC conditioning regimens), and hence the choice of comparator (RIC Bu/Flu) means that the relative effectiveness and cost effectiveness of Treo/Flu has not been evaluated against the range of comparator regimens that would be relevant for the full population defined in the scope.

The CS and response to clarification questions included two registry studies, European Society for Blood and Marrow Transplantation (EBMT) and US Center for International Blood and Marrow Transplant Research (CIBMTR), which were commissioned by the company in order to provide information about the comparative effectiveness of Treo/Flu versus other conditioning regimens. Both studies included only registry patients who matched the reported criteria used in MC-FludT.14/L Trial II to define patients at increased risk and not eligible for standard high-intensity MAC conditioning regimens, however, the conditioning regimens received by registry patients in these studies included some high-intensity MAC regimens. The ERG therefore questions whether MC-FludT.14/L Trial II used any additional criteria to define increased risk patients, not eligible for standard high-intensity MAC conditioning regimens. Furthermore, the ERG is unclear as to how similar the definition of patients at increased risk for standard conditioning therapies (i.e. not eligible for standard high-intensity MAC), used in MC-FludT.14/L Trial II, is to any such definition that would be generally applied in practice; it is important to establish a consistent definition in order to inform any recommendations for this population.

Irrespective of whether/the extent to which the registry studies can provide comparative effectiveness data for Treo/Flu versus relevant comparator regimens (including standard high-intensity MAC regimens where applicable), data from these studies have not been utilised in the cost effectiveness modelling. The ERG therefore considers that the evidence included in the submission is sufficient to support an assessment of the cost effectiveness of Treo/Flu versus RIC Bu/Flu in adults with AML or MDS, who are at increased risk for standard conditioning therapies (i.e. not eligible for standard high-intensity MAC regimens). However, the evidence included in the submission is not sufficient to support an assessment of the cost effectiveness of Treo/Flu for the full scope population or versus any of the other comparators defined in the scope.

Table 1.1: Efficacy results of MC-FludT.14/L Trial II

	Treosulfan (10 g/m²/day) + Fludarabine (30 g/m²/day)	**Busulfan (3.2 mg/kg/day) + Fludarabine (30 g/m²/day)**
Number randomised	280	290
Number analysed (FAS: patients who received conditioning treatment and HSCT)	268	283
Median follow-up[a], months (range of those surviving)	29.7 (3.0 to 52.1)	29.4 (3.0 to 54.3)
Primary outcome – Event-free survival (EFS) within 24 months after alloHSCT		
Patients with event	97 (36.2%)	137 (48.4%)
Death[b]	35 (13.1%)	56 (19.8%)
Relapse/progression[b]	61 (22.8%)	72 (25.4%)
Primary graft failure[b]	1 (0.4%)	1 (0.4%)
Secondary graft failure[b]	0 (0.0%)	8 (2.8%)
Event-free survival at 12 months[c] [%] (95% CI)	70.0 (64.1 to 75.1)	60.8 (54.9 to 66.3)
Event-free survival at 24 months[c] [%] (95% CI)	65.7 (59.5 to 71.2)	51.2 (45.0 to 57.0)
Event-free survival at 36 months[c] [%] (95% CI)	[redacted]	[redacted]
Hazard ratio [HR][d] (95% CI)	0.64 (0.49 to 0.84)	
Secondary outcome – Overall survival (OS) within 24 months after alloHSCT		
Patients with event	81 (30.2%)	112 (39.6%)
Overall survival at 12 months[c] [%] (95% CI)	77.8 (72.3 to 82.3)	71.8 (66.1 to 76.7)
Overall survival at 24 months[c] [%] (95% CI)	72.7 (66.8 to 77.8)	60.2 (54.0 to 65.8)
Overall survival at 36 months[c] [%] (95% CI)	[redacted]	[redacted]
HR[d] (95% CI)	0.64 (0.48 to 0.87)	
Secondary outcome – Cumulative incidence of relapse/progression 24 months after HSCT		
Patients with event	61 (22.8%)	72 (25.4%)
Patients without event (censored) or with competing event	207 (77.2%)	211 (74.6%)

	Treosulfan (10 g/m²/day) + Fludarabine (30 g/m²/day)	Busulfan (3.2 mg/kg/day) + Fludarabine (30 g/m²/day)
Censored	171 (63.8%)	146 (51.6%)
Death[b]	35 (13.1%)	56 (19.8%)
Primary graft failure[b]	1 (0.4%)	1 (0.4%)
Secondary graft failure[b]	0 (0.0%)	8 (2.8%)
Cumulative incidence at 12 months [%] (95% CI)	19.1 (14.4 to 23.8)	21.7 (16.9 to 26.5)
Cumulative incidence at 24 months [%] (95% CI)	22.0 (16.9 to 27.1)	25.2 (20.0 to 30.3)
Cumulative incidence at 36 months [%] (95% CI)	[redacted]	[redacted]
HR[e] (95% CI)	0.82 (0.59 to 1.16)	
Secondary outcome - engraftment		
Primary graft failure[f]	1/268 (0.4%)	1/283 (0.4%)
Secondary graft failure[f]	0/263 (0.0%)	8/279 (2.9%)
Secondary outcome (not specified in scope) – Cumulative incidence of non-relapse mortality (NRM) 24 months after HSCT		
Patients with event	35 (13.1%)	56 (19.8%)
Patients without event (censored) or with competing event	233 (86.9%)	227 (80.2%)
Censored	171 (63.8%)	146 (51.6%)
Relapse/Progression[b]	61 (22.8%)	72 (25.4%)
Primary Graft Failure[b]	1 (0.4%)	1 (0.4%)
Secondary Graft Failure[b]	0 (0.0%)	8 (2.8%)
Cumulative incidence at 12 months [%] (95% CI)	10.5 (6.8 to 14.2)	14.3 (10.2 to 18.4)
Cumulative incidence at 24 months [%] (95% CI)	12.0 (8.0 to 15.9)	20.4 (15.5 to 25.2)
Cumulative incidence at 36 months [%] (95% CI)	[redacted]	[redacted]
HR[g] (95% CI)	0.63 (0.41 to 0.97)	
Secondary outcome (not specified in scope) – Transplantation-related mortality (TRM)		

	Treosulfan (10 g/m²/day) + Fludarabine (30 g/m²/day)	**Busulfan (3.2 mg/kg/day) + Fludarabine (30 g/m²/day)**
Patients with event	33 (12.3%)	58 (20.5%)
Patients without event	235 (87.7%)	225 (79.5%)
Transplantation-related mortality at 12 months[c] [%] (95% CI)	11.7 (8.3 to 16.3)	16.2 (12.2 to 21.3)
Transplantation-related mortality at 24 months[c] [%] (95% CI)	12.8 (9.2 to 17.7)	24.1 (19.1 to 30.2)
HR[d] (95% CI)	0.52 (0.34 to 0.82)	

Based on Tables 18, 19, 20, 21 and 23 of the CS

[a] Based on reverse Kaplan-Meier estimates for overall survival; [b] Only if this event occurred first; [c] Based on Kaplan-Meier estimates; [d] Adjusted for donor type as factor, and risk group and centre as strata using Cox regression model; [e] Adjusted for donor type as factor and risk group as stratum using Fine and Gray model; [f] Rate of primary/secondary graft failure calculated as number of patients with graft failure by the number of patients at risk; [g] Adjusted for donor type as factor

alloHSCT = allogeneic haematopoietic stem cell transplantation; CI = confidence interval; CS = company submission; EFS = event-free survival; FAS = full analysis set; HR = hazard ratio; HSCT = Haematopoietic stem cell transplantation; NRM = non-relapse mortality; OS = Overall survival; TRM = transplantation-related mortality

Table 1.2: Summary of treatment emergent adverse events MC FludT.14/L

Safety analysis set (SAS)	Busulfan (N=283)	Treosulfan (N=270)	Total (N=553)
Any AE, n (%)			
Patients with any AE	272 (96.1)	250 (92.6)	522 (94.4)
Patients with AEs of at least CTCAE Grade III	151 (53.4)	148 (54.8)	299 (54.1)
Drug-related AEs, n (%)			
Patients with any drug-related AEs	192 (67.8)	170 (63.0)	362 (65.5)
Patients with drug-related AEs of at least CTCAE Grade III	82 (29.0)	72 (26.7)	150 (27.8)
SAE, n (%)			
Patients with any SAE	20 (7.1)	23 (8.5)	43 (7.8)
Resulting in death	6 (2.1)	8 (3.0)	14 (2.5)
Life-threatening	8 (2.8)	13 (4.8)	21 (3.8)
Hospitalisation or prolongation of hospitalisation	9 (3.2)	8 (3.0)	17 (3.1)
Drug-related SAEs, n (%)			
Patients with any drug related SAE	9 (3.2%)	9 (3.3%)	18 (3.3%)
Maximum CTCAE grade of adverse events [n (%)]			
Patients with AEs of a maximum CTCAE grade I	46 (16.3%)	41 (15.2%)	87 (15.7%)
Patients with AEs of a maximum CTCAE grade II	75 (26.5%)	61 (22.6%)	136 (24.6%)
Patients with AEs of a maximum CTCAE grade III	134 (47.3%)	123 (45.6%)	257 (46.5%)
Patients with AEs of a maximum CTCAE grade IV	14 (4.9%)	18 (6.7%)	32 (5.8%)
Patients with AEs of a maximum CTCAE grade V	3 (1.1%)	7 (2.6%)	10 (1.8%)

Based on Table 30 of the CS
AE = adverse event; CS = company submission; CTCAE = Common Terminology Criteria for Adverse Events; N = number of participants; n = number of patients in category; SAE = Serious adverse event; SAS = safety analysis set

1.3 Summary of the key issues in the cost effectiveness evidence

A range of resources were searched. Searches were well documented, making them transparent and reproducible. Subject headings and free text terms were applied and combined correctly for each facet. Searches for interventions and comparators would have benefited from the inclusion of more synonyms and the use of validated search filters would have improved the sensitivity of searches for economic, cost and utility studies.

Separate economic and health utilities searches were undertaken in the Embase database. Searches for economic and health utilities studies in PubMed and the Cochrane Library were combined with searches for RCTs. Conference proceedings were also searched for cost effectiveness studies. The CS provided sufficient details for the ERG to appraise the searches.

The use of validated and tested cost and utility search filters would have improved the comprehensiveness and sensitivity of these searches.

The modelling approach considered by the company is in line with those in two recent NICE technology appraisals on AML (TA523 and TA552). The ERG considers this approach appropriate for the decision problem at hand.

The ERG conducted a more detailed overall goodness-of-fit assessment based on all the information presented by the company either in the main submission document, in the appendices or in the response to the request for clarification. Based on this assessment, the ERG was able to identify the best candidate distributions to model overall survival (OS) and event-free survival (EFS), which in some cases did not match the ones selected by the company, and to define alternative scenario analyses.

Justification for adverse events, resource use, costs and utility sources chosen by the company was clear and the ERG agreed with the company's choice in relation to the application of costs and utilities in the model. Therefore, the ERG did not amend the costs or utility parameters chosen by the company for the base-case. The robustness of the model results to changes in both cost and utility parameters was tested with scenario analyses, conducted by both the company and the ERG.

1.4 Summary of the ERG's preferred assumptions and resulting ICER

The ERG preferred changes to the company base-case are described in Section 7.1.2 and summarised below:

1. Correcting OS and EFS implementation.
2. Using a rescaling factor (year to day) equal to 1/365.25 (instead 1/364).
3. Modelling OS according to a non-mixture-cure model (NMCM) Weibull distribution.
4. Using most recent UK life tables.

The cost effectiveness results of the ERG preferred base-case are presented in Table 1.3. The implementation of the ERG preferred assumptions resulted in treosulfan generating 0.78 more QALYs than busulfan at lower costs (-£17,689). Therefore, treosulfan dominated busulfan as in the company base-case. In the ERG base-case, both cost savings and QALY gains for treosulfan were smaller than in the company base-case. The assumption with the largest impact on the incremental costs was correcting the calculations of the overall and event-free survival probabilities. This results in incremental costs increased by £9,176 and incremental QALYs decreased by 0.05. The assumption with the largest impact on the incremental QALYs was modelling OS according to a NMCM Weibull distribution. This results in incremental QALYs decreased by 0.06 and incremental costs decreased by £3,151. The other two changes made by the ERG had a minor impact on the results.

Table 1.3: ERG preferred deterministic base-case results (discounted)

	Total costs (£)	Total LYGs	Total QALYs	Incr. costs (£)	Incr. LYGs	Incr. QALYs	ICER versus baseline (£/QALY)
Treosulfan	£128,147	8.75	6.49	-£17,689	0.92	0.78	Treosulfan dominates
Busulfan	£145,836	7.84	5.71				

ERG = Evidence Review Group; ICER = incremental cost effectiveness ratio, Incr. = incremental, LYG = life year gained, QALY = quality-adjusted life year

1.5 Summary of exploratory and sensitivity analyses undertaken by the ERG

The ERG also conducted a probabilistic sensitivity analyses (PSA) using the ERG preferred base-case assumptions. The probabilistic results were in line with the ERG and company base-case since treosulfan dominated busulfan.

Several additional scenarios were conducted by the ERG. Assumptions regarding the selection of parametric curves for EFS and OS, the duration of the "cure point", alternative approaches to long-term mortality or different standardised mortality ratios (SMRs) to model HSCT mortality should have been explored by the company but were missing in the company submission. Scenarios using alternative assumptions on costs, different health state utilities and AE disutilities were also conducted by the ERG. The results of these analyses confirmed the ERG expectations since in all the scenarios explored, treosulfan was dominant.

Finally, the ERG also presented results for AML and MDS patients separately. In these analyses, the company assumed the same EFS and OS distributions as in the base-case with the pooled population. However, the ERG considered that the selection of survival models should be based on subgroup-specific data. Based on these analyses, the ERG selected OS and EFS parametric curves that were different than the ones selected by the company. Furthermore, all subgroup analyses were run selecting subgroup-specific patient characteristics (instead of using the characteristics of the pooled population). In both subgroups, treosulfan was dominant except in one scenario for the MDS population where it was nevertheless cost effective (the ICER was £8,171). In the AML subgroup, cost savings for treosulfan were larger than in the MDS subgroup, but with smaller QALY gains. The uncertainty in the MDS subgroup seems to be larger too, given the overall poor fit of the OS and specially EFS curves.

2. Background

2.1 Introduction

In this report, the Evidence Review Group (ERG) provides a review of the provided evidence submitted by medac GmbH in support of treosulfan, trade name Trecondi®, with fludarabine for treating malignant disease in remission before allogeneic haematopoietic stem cell transplantation (HSCT) in adults, children and young people.

2.2 Background

HSCT is a potentially curative therapy for many cancers and non-malignant disorders.[1]

Autologous HSCT uses the patient's own cells, allogeneic HSCT uses cells from a donor.[2] Both HSCTs are potentially curative for a host of different diseases, typically for a specific disease autologous or allogeneic will be the more successful treatment.[2] *"Allogeneic HSCT is recommended for acute and chronic leukaemias, myelo-dysplastic syndromes (MDS), HL* [Hodgkin lymphoma]*, NHL* [Non-Hodgkin lymphoma]*, and MM* [multiple myeloma]*".*[2]

However, allogeneic HSCT involves a certainty of rejection unless the patient is first treated with myeloablative conditioning (MAC). Standard MAC regimens generally lead to low relapse rates, but have high treatment-related toxicity and transplant-related mortality (TRM).[3]

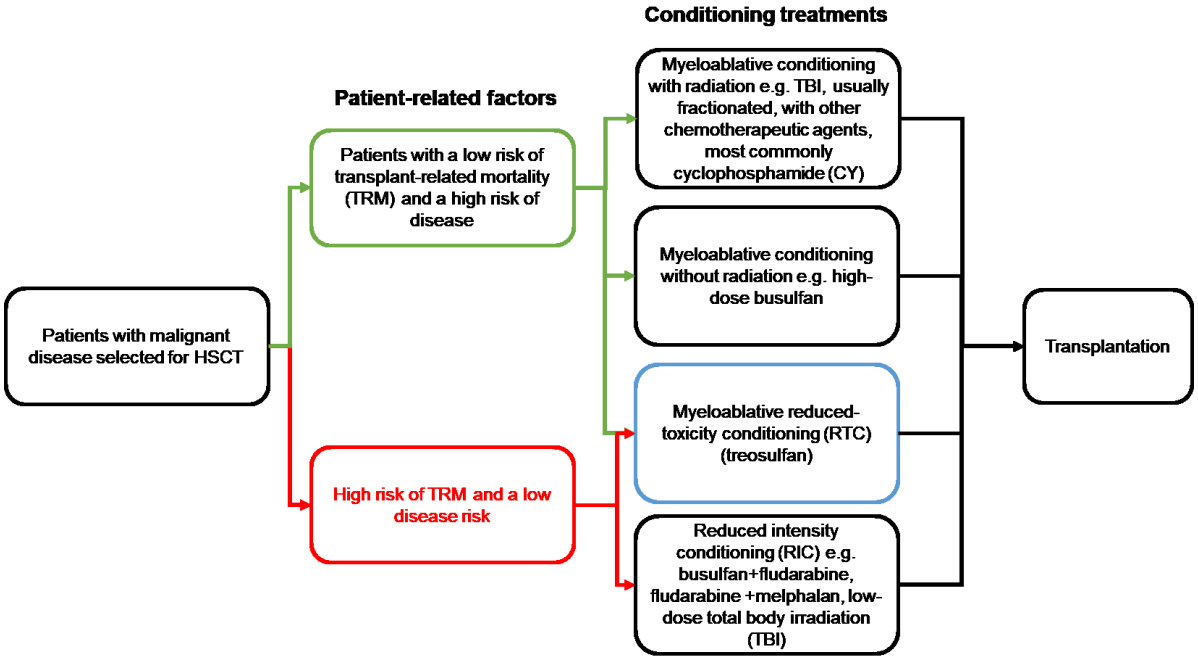
Non-myeloablative (NMA) conditioning and reduced intensity conditioning (RIC) was developed for patients such as the elderly and those with comorbidities where myeloablative conditioning is not considered optimal and to minimise treatment-related toxicity, non-relapse mortality (NRM) and TRM.[1] However, lower dose intensity conditioning has higher rate of relapse after allogeneic HSCT.[3]

The company submission (CS) states that 1,594 patients received allogeneic HSCT in the United Kingdom (UK) (39% of all HSCTs).[4]

According to the CS, allogeneic HSCT is increasingly used in non-malignant diseases, at least for a number of non-malignant diseases where autologous HSCT is not recommended.[1, 2]

The CS states that the European Society for Blood and Marrow Transplantation (EBMT) advises focussing on the concept that patients with a high risk for TRM and a low disease risk should receive a different conditioning regimen from patients with a low risk for TRM and high risk disease.[1, 5] To that effect, the CS is suggesting treosulfan with fludarabine as a conditioning treatment for some patient combinations of TRM risk and a disease risk (see Figure 2.1).

Figure 2.1: Company suggested clinical context of treosulfan with fludarabine in conditioning treatment for malignant haematological diseases selected for HSCT



Based on Figure 2 of the CS[1]
CS = company submission; Cy = cyclophosphamide; HSCT = haematopoietic stem cell transplantation; RTC = reduced toxicity conditioning; TBI = total body irradiation; TRM = transplantation-related mortality

Standard intensity conditioning carries a relatively high risk of mortality and morbidity but aims to reduce toxicity while maintaining efficacy of engraftment and anti-tumour effect.[1] NMA and RIC are used to expand the patient populations able to receive allogeneic HSCT. However, the relapse rate for NMA and RIC is increased compared to standard intensity conditioning.[1]

The CS states that a medac clinical trial found good tolerability of treosulfan 4.7-times above the maximum tolerated dose (MTD) with diarrhoea, mucositis/stomatitis, and metabolic acidosis being the dose-limiting toxicities.[1] However, severe liver, renal, neural or lung toxicities as otherwise observed with high-dose alkylating agents were not observed after high-dose treosulfan.[1, 6] The CS states that *'as a single agent, grade III/IV toxicities are very rarely observed'* and that *'treosulfan lacks hepatic and renal toxicity and only very rarely causes lung toxicity'*.[1, 7]

2.3 Critique of company's description of underlying health problem.

Overall, the ERG agrees with the description of the underlying health problem. However, as detailed in Section 3, the ERG has a number of comments on how this was implemented in the decision problem.

3. Critique of company's definition of decision problem

Table 3.1: Statement of the decision problem (as presented by the company)

	Final scope issued by NICE/ reference case	Decision problem addressed in the company submission	Rationale if different from the final NICE scope	ERG comment
Population	Adults, children and young people with malignant disease that is in remission before allogenic haematopoietic stem cell transplantation	As final scope		The population addressed in the CS is narrower than in the final scope, see Section 3.1
Intervention	Treosulfan with fludarabine	As final scope		As final scope
Comparator(s)	Conditioning treatments (either high dose or reduced intensity): • cyclophosphamide and total body irradiation • cyclophosphamide and busulfan • busulfan with fludarabine • established clinical management without treosulfan with fludarabine.	As final scope		The CS reports results on one comparator in a defined population but provides insufficient evidence for the other comparators, see Section 3.3
Outcomes	The outcome measures to be considered include: • overall survival • event-free survival • rates of relapse • success of stem cell transplantation (engraftment) • adverse effects of treatment • health-related quality of life	As final scope, with the addition of • non-relapse mortality (NRM).	NRM has been added as a significant reduction in NRM is the reason for the overall survival benefit and event free survival benefit observed with treosulfan-based conditioning.	Health-related quality of life was not reported in the CS, see Section 3.4

	Final scope issued by NICE/ reference case	Decision problem addressed in the company submission	Rationale if different from the final NICE scope	ERG comment
Perspective for outcomes		All direct health effects for patients were considered.		
Perspective for costs	The reference case stipulates that the cost effectiveness of treatments should be expressed in terms of incremental cost per quality-adjusted life year.	The economic analysis was conducted from a National Health Service (NHS) and Personal Social Services (PSS) perspective, as per the NICE reference case.		
Time horizon	The reference case stipulates that the time horizon for estimating clinical and cost effectiveness should be sufficiently long to reflect any differences in costs or outcomes between the technologies being compared.	A lifetime (40 year) time horizon was assumed as HSCT is a potentially curative treatment. Shorter time horizons of 5 and 10 years were considered in scenario analysis.		
Synthesis of evidence on health effects		The systematic literature review identified a cost-effectiveness analysis of two reduced intensity conditioning (RIC) regimens; this was the sole cost-effectiveness analysis relating to HSCT conditioning therapy. Due to a lack of HSCT conditioning regimen cost-utility studies, economic analyses in AML and MDS were searched for in a targeted literature review.	Cost utility and quality-of-life (QoL) data were not collected in the pivotal phase III trial.	
Measuring and valuing health effects	Health effects should be expressed in QALYs. The EQ-5D is the preferred measure of health-related quality of life in adults.	Health effects are expressed in QALYs. Quality-of life (QoL) data were not collected during the phase III trial. Appropriate published QoL data, as identified in the systematic literature review and targeted	QoL data were not collected during the phase III trial and no appropriate sources of UK EQ-5D data were identified. Therefore, mapping had to be used.	

	Final scope issued by NICE/ reference case	Decision problem addressed in the company submission	Rationale if different from the final NICE scope	ERG comment
		literature search was utilised. No appropriate sources of UK EQ-5D data were identified. Mapping was used to convert data from other measures into EQ-5D utility values where required. Utility values obtained from TTO and DCE studies were included as a scenario.		
Source of data for measurement of health-related quality of life	Reported directly by patients and/or carers.	Published sources of patient reported health state utility values (HSUVs) were identified through systematic literature review, as well as review of prior AML and MDS technology appraisals and targeted searching of the literature.	HRQoL data were not collected during the phase III trial. Appropriate published patient reported HRQoL data, as identified in the systematic literature review and targeted literature search was utilised and is presented as QALYs.	
Source of preference data for valuation of changes in health-related quality of life	Representative sample of the UK population	No UK EQ-5D data could be identified. Mapping was used to convert data from other measures into EQ-5D utility values where required. Where possible mapping algorithms generated from UK samples were used.	HRQoL data were not collected during the phase III trial and no UK EQ-5D data could be identified. Mapping was used to convert data from other measures into EQ-5D utility values where required. Where possible mapping algorithms generated from UK samples were used.	
Equity considerations	An additional QALY has the same weight regardless of the other characteristics of the individuals receiving the health benefit.	There are no equity considerations with treosulfan-based conditioning.		

	Final scope issued by NICE/ reference case	Decision problem addressed in the company submission	Rationale if different from the final NICE scope	ERG comment
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.	As reference case.		
Discounting	The same annual rate for both costs and health effects (currently 3.5%).	As reference case.		

Based on Table 1 of the CS[1]

AML = acute myeloid leukaemia; CS = company submission; DCE = discrete choice experiment; EQ-5D = European Quality of Life-5 Dimensions; ERG = Evidence Review Group; HRQoL = health-related quality of life; HSCT = haematopoietic stem cell transplantation; HSUV = health state utility value; MDS = myelodysplastic syndrome; NHS = National Health Service; NICE = National Institute for Health and Care Excellence; NRM = non-relapse mortality; PSS = Personal Social Services; QALY = quality-adjusted life year; RIC = reduced intensity conditioning; TTO = time trade-off; UK = United Kingdom

3.1 Population

The population in the submission is a subgroup of that defined in the scope and covered in the licenced indication for treosulfan.[8, 9]

The decision problem, described by the company in the CS, states that: *"This technology appraisal evaluates the clinical and cost-effectiveness of treosulfan as a conditioning treatment for malignant disease prior to allogeneic haematopoietic stem cell transplantation (alloHSCT) in paediatric and adult patients older than one month."*

However, the submission relies, primarily, on one randomised controlled trial (RCT) of treosulfan (MC-FludT.14/L Trial II) assessed the efficacy and safety of treosulfan in combination with fludarabine for conditioning therapy as part of allogeneic haematopoietic stem cell transplantation (alloHSCT) in adults with AML or MDS who were at increased risk for standard conditioning therapies (i.e. not eligible for standard MAC busulfan- or total body irradiation (TBI)-based regimens).[10] Patients at increased risk for standard conditioning therapies were defined as those who were patients aged ≥50 years at transplant and/or who had a haematopoietic cell transplantation co-morbidity index (HCT-CI) score >2.[10]

ERG comment: The company were asked to clarify whether the intended population for this submission is as described in the scope or the more restricted population, which is consistent with the evidence that has been presented.[11] The following response was provided[12]:

"We have presented in our submission the best available evidence for treosulfan and this is based on the pivotal phase III trial (MC-FludT.14/L Trial II). This trial, with 570 randomised patients, was the largest ever prospective randomised clinical trial (RCT) comparing two conditioning regimens.[10] This trial focusses on "patients who cannot tolerate such MAC regimens and would otherwise be treated with less efficacious RIC regimens i.e. patients aged ≥ 50 years at transplant and/or with a Haematopoietic Cell Transplantation-Comorbidity Index (HCT-CI) score >2 (indicating an intermediate to high risk of co-morbidities which influence NRM)."

However, we believe the evidence from this population is applicable to the broader population of "Adults, children and young people with malignant disease that is in remission before allogenic haematopoietic stem cell transplantation" in line with the final scope. In Appendix L we provided the "Haematologists' Position Paper" from Prof. Suttorp et al. This position paper includes the opinion of six European KOLs (including 2 UK KOLs) and states:

"The signatories of this consensus statement wish to share their opinions and strong belief that TREO-based conditioning should not only be reserved to treat defined malignancies such as AML or MDS in certain age groups. We consider the existing trial data as compelling to support the use of TREO in most patients with malignant diseases requiring treatment with myeloablative conditioning followed by alloHSCT."

Therefore, we feel justified in believing that the trial population can be broadened in line with the Final Scope."

ERG comment: The ERG remains very concerned that population included in MC-FludT.14/L Trial II is not representative of the full population defined in the scope. Specifically, there is a lack of evidence about the effectiveness of treosulfan/fludarabine conditioning regimens in people who are able to tolerate standard MAC regimens, in adults with malignancies other than AML or MDS, and in children.

3.2 Intervention

The CS (Section B.1.2) includes the following statement[1]:

"Treosulfan (Trecondi ®, medac GmbH) is a water-soluble prodrug of a bifunctional alkylating agent. Due to its proven antileukaemic and immunosuppressive activity, treosulfan in combination with fludarabine and other agents has been developed as conditioning regimen prior to alloHSCT in adults and children with malignant disease.

Treosulfan is administered by intravenous infusion which should be supervised by a physician experienced in conditioning treatment and alloHSCT. Monitoring procedures outlined in the SmPC (see Appendix C) should be followed."

The key product characteristics are summarised in Table 3.2.

Table 3.2: Treosulfan product characteristics

UK approved name and brand name	**Treosulfan (Brand name Trecondi®)**
Mechanism of action	Treosulfan is a prodrug of a bifunctional alkylating agent with cytotoxic activity to haematopoietic precursor cells. The activity of treosulfan is due to the spontaneous conversion into a mono-epoxide intermediate and L-diepoxybutan. The epoxides formed alkylate nucleophilic centres of deoxyribonucleic acid (DNA) and are able to induce DNA cross-links which are considered responsible for the stem cell depleting and antineoplastic effects.[13]
Marketing authorisation/CE mark status	In December 2017, medac submitted a marketing authorisation application (EMEA/H/C/004751) for treosulfan in a centralised procedure, according to Regulation (EC) No. 726/2004. On 13 December 2018, the Committee for Medicinal Products for Human Use (CHMP) adopted a positive opinion, recommending the granting of a marketing authorisation for the medicinal product treosulfan, intended for the conditioning treatment prior to alloHSCT. Medac anticipate that the final European Commission decision will not be published until May 2019 with the European Public Assessment Report (EPAR) published shortly afterwards. An appeal was submitted in March 2019 regarding the orphan status of treosulfan. [redacted]. An injected form of treosulfan has had marketing authorisation in the UK since January 1992 and is indicated for the palliative treatment of epithelial ovarian cancer.[14]
Indications and any restriction(s) as described in the summary of product characteristics (SmPC)	The anticipated full indication is "Treosulfan in combination with fludarabine is indicated as part of conditioning treatment prior to allogeneic haematopoietic stem cell transplantation (alloHSCT) in adult patients with malignant and non-malignant diseases, and in paediatric patients older than one month with malignant diseases."
Method of administration and dosage	Administration of treosulfan should be supervised by a physician experienced in conditioning treatment followed by alloHSCT.[13]

<table>
<tr><th>UK approved name and brand name</th><th>Treosulfan (Brand name Trecondi®)</th></tr>
<tr><td></td><td>
Adults with malignant disease: Treosulfan is given in combination with fludarabine. The recommended dose and schedule of administration is as follows:
Treosulfan 10 g/m² body surface area (BSA) per day as two-hour i.v. infusion, given on three consecutive days (day -4, -3, -2) before stem cell infusion (day 0). The total treosulfan dose is 30 g/m²
Fludarabine 30 mg/m² BSA per day as a 0.5-hour i.v. infusion, given on five consecutive days (day -6, -5, -4, -3, -2) before stem cell infusion (day 0). The total fludarabine dose is 150 mg/m²
Treosulfan should be administered before fludarabine on days -4, -3, -2 (FT_{10} regimen).
Paediatric population: Treosulfan is given in combination with fludarabine, with thiotepa (intensified regimen; $FT_{10-14}TT$ regimen) or without thiotepa (FT_{10-14} regimen). The recommended dose and schedule of administration is as follows:
Treosulfan 10–14 g/m² BSA/day as two-hour i.v. infusion, given on three consecutive days (day -6, -5, -4) before stem cell infusion (day 0). The total treosulfan dose is 30–42 g/m². The dose of treosulfan should be adapted to the patient's BSA (see table below).
Treosulfan dose based on patient BSA.
<table>
<tr><th>Body surface area (m²)</th><th>Treosulfan dose (g/m²)</th></tr>
<tr><td>≤0.5</td><td>10.0</td></tr>
<tr><td>>0.5–1.0</td><td>12.0</td></tr>
<tr><td>>1.0</td><td>14.0</td></tr>
</table>
Fludarabine 30 mg/m² BSA/day as a 0.5-hour i.v. infusion, given on five consecutive days (day -7, -6, -5, -4, -3) before stem cell infusion (day 0). The total fludarabine dose is 150 mg/m².
Treosulfan should be administered before fludarabine.
Thiotepa (intensified regimen) 2 × 5 mg/kg, given as two i.v. infusions over 2-4 hours on day -2 before stem cell infusion (day 0).
</td></tr>
<tr><td>Additional tests or investigations</td><td>No additional tests or investigations are needed for treatment eligibility outside of those required in clinical practice for patients with malignant disease requiring chemotherapy.
After initiating treosulfan, patients would require no tests or investigations additional to those that would already be performed following treatment with standard intensive chemotherapy.</td></tr>
<tr><td>List price and average cost of a course of treatment</td><td>Treosulfan is priced at £53.83 (1 g vial) and £208.03 (5 g vial).
Total treatment cost for treosulfan is £2,496.41 (including wastage).
Fludarabine's list price is £147.07 (50 mg powder for solution for injection vials) and the total treatment cost for fludarabine is £3,059 (including wastage).</td></tr>
<tr><td>Patient access scheme (if applicable)</td><td>Not applicable.</td></tr>
<tr><td colspan="2">Based on Table 2 of the CS.[1]</td></tr>
</table>

UK approved name and brand name	Treosulfan (Brand name Trecondi®)
alloHSCT = allogeneic haematopoietic stem cell transplantation; BSA = body surface area; CE = Conformité Européenne (engl. European conformity); CHMP = Committee for Medicinal Products for Human Use; CS = company submission; DNA = deoxyribonucleic acid; EPAR = European Public Assessment Report; g = gram; mg = milligram; SmPC = Summary of product characteristics; UK = United Kingdom	

ERG comment: In line with the NICE scope, the CS considered treosulfan in combination with fludarabine. The dosing schedule used in the included RCT (MC-FludT.14/L Trial II)[10] is consistent with the recommended dosing schedule, as described in Table 3.2.

3.3 Comparators

The NICE scope lists the comparators as[8]:

- Standard high-intensity (myeloablative) conditioning regimens
    - cyclophosphamide and total body irradiation
    - cyclophosphamide and busulfan
    - cyclophosphamide and thiotepa
    - high-dose busulfan with fludarabine with or without thiotepa
- Reduced intensity conditioning regimens
    - low-dose busulfan with fludarabine
    - melphalan plus fludarabine

In its definition of the decision problem (Table 3.1), the company indicates that the comparator(s) are *"as final scope"*, however, the comparators listed in Table 3.1 do not include regimens with thiotepa and the reduced intensity conditioning regimen, melphalan plus fludarabine, is also not explicitly listed.[1]

The included RCT, MC-FludT.14/L Trial II,[10] only provides comparative efficacy data for treosulfan in combination with fludarabine versus one alternative conditioning therapy, a RIC regimen of busulfan in combination with fludarabine. The company were asked to confirm whether the analysis of registry data (described in Section B.2.2.4 of the CS) represented the only data being submitted on the comparative effectiveness of treosulfan with fludarabine vs. conditioning regimens other than busulfan in combination with fludarabine.[11] The following response was provided[12]:

"Our submission focusses on reduced intensity conditioning because this is where we have direct phase III clinical trial evidence and also where there is an unmet need for a large proportion of patients. Further evidence for the comparison of treosulfan with fludarabine against cyclophosphamide, high-dose busulfan with thiotepa and melphalan plus fludarabine is based on the registry analysis which medac commissioned. The European Society for Blood and Marrow Transplantation (EBMT) registry was analysed and the results were discussed in section B.2.2.2.4 of the submission (a PDF is now provided as requested in question D2).[15]

In addition, the US Center for International Blood and Marrow Transplant Research (CIBMTR) was also recently asked by medac to analyse their registry and compare it to the data from the MC-FludT.14/L Trial II. The CIBMTR report[16] provides a comparison of allogeneic transplantation outcomes for acute myeloid leukemia (AML) and myelodysplastic syndrome (MDS) with full or reduced intensity busulfan with fludarabine regimens and busulfan-cyclophosphamide regimen in the United States to the results of the MC-FludT.14/L Trial II.

Lastly we refer you to the Haematologists' Position Paper within Appendix L in which six European Key Opinion Leaders summarise that "TREO/FLU conditioning leads to better outcomes than busulfan-cyclophosphamide, as well as melphalan-based regimens."

The company's response and details of the two registry studies are discussed further in Section 4.2 of this report.

ERG comment: The ERG notes that the registry studies described do not provide data for the comparative effectiveness of treosulfan plus fludarabine (Treo/Flu) versus the full range of comparators in the whole population, as specified in the final NICE scope.[8] The ERG therefore considers that the evidence included in the submission is sufficient to support an assessment of the cost effectiveness of Treo/Flu versus RIC Bu/Flu in adults with AML or MDS, who are at increased risk for standard conditioning therapies (i.e. not eligible for standard high-intensity MAC regimens). However, the evidence included in the submission is not sufficient to support an assessment of the cost effectiveness of Treo/Flu for the full scope population or versus any of the other comparators defined in the scope.

3.4 Outcomes

The NICE scope lists the following outcome measures[8]:

- overall survival
- event-free survival
- relapse
- success of stem cell transplantation (engraftment)
- adverse effects of treatment
- health-related quality of life

The company's definition of the decision problem (Table 3.1) includes the additional, relevant outcome non-relapse mortality (NRM).

ERG comment: The ERG notes that the included RCT (MC-FludT.14/L Trial II) did not assess any health-related quality of life (HRQoL) outcomes and no other studies reporting HRQoL outcomes were included in the clinical effectiveness section of the CS.[1, 10]

3.5 Other relevant factors

The CS (Section B.1.4) states: *"We do not envisage any equity or equality issues with the use of treosulfan in combination with fludarabine as conditioning treatment prior to alloHSCT in adults and children with malignant disease."*

ERG comment: The ERG has no further comments on other factors.

4. Clinical effectiveness

4.1 Critique of the methods of review(s)

The company conducted a systematic review to *"identify publications reporting on the clinical efficacy and safety, health-related quality of life (HRQoL), resource use/costs and cost effectiveness of treosulfan in combination with fludarabine in patients with AML, ALL, MDS or MM and undergoing HSCT"*.[17] This section of the ERG report critiques the methods of the review including searching, inclusion criteria, data extraction, quality assessment and evidence synthesis.

The systematic review was described, in detail, in Appendix D of the CS.[17]

4.1.1 Searches

Appendix A of the CS details a systematic search of the literature used to identify clinical effectiveness literature undertaken on 11 and 12 February 2019.[1] A summary of the sources searched is provided in Table 4.1.

Table 4.1: Data sources for the clinical effectiveness systematic review

Search strategy element	Resource	Host/Source	Date Range	Date searched
Electronic databases	Embase	Embase.com	2008-2019	11 February 2019. Update searches on 20 June 2019
	PubMed	www.ncbi.nlm.nih.gov/pubmed/	2008-2019	12 February 2019
	Cochrane Library (Cochrane Reviews, Cochrane Protocols, Clinical answers, Editorials and Special Collections)	http://onlinelibrary.wiley.com/cochranelibrary/search/	2008-2019	12 February 2019
Conference proceedings	ASH	2017 – host not reported 2018 - www.bloodjournal.org/page/ash-annual-meeting-abstracts?sso-checked=true	2017-2018	Not reported
	BSBMT	2017 - www.bsbmt.org/wp-content/uploads/2017/08/BSBMT-News-18d.pdf 2018 – not reported	2017-2018	Not reported
	CIBMTR/ ASBMT	2017 - www.bloodjournal.org/content/130/suppl_1?sso-checked=true	2017-2018	Not reported

Search strategy element	Resource	Host/Source	Date Range	Date searched
		2018 – not reported		
	EHA	2017 – https://ehaweb.org/assets/Uploads/EHA22-Abstract-Book.pdf 2018 - https://journals.lww.com/hemasphere/Citation/2018/06001/23rd_Congress_of_the_European_Hematology.1.aspx	2017-2018	Not reported
	EBMT	https://www.ebmt.org/annual-meeting>abstracts	2017-2018	Not reported
ASBMT = American Society for Blood and Marrow Transplantation; ASH = American Society of Haematology; BSBMT = British Society of Blood and Marrow Transplantation; CIBMTR = Centre for International Blood and Marrow Transplant Research; EBMT = European Society for Blood and Marrow Transplantation; EHA = European Hematology Association				

ERG comment:

- The selection of databases searched was comprehensive, and searches were clearly reported and reproducible. The database name, host and date searched were provided. A range of resources additional to databases were searched to identify further relevant studies.
- Some additional synonyms for drug names were not included in the search. For example, Embase searches did not include the term "ovastat" which is a brand name for treosulfan; Embase, PubMed and Cochrane searches did not include "oforta" which is a brand name for fludarabine or "evomela" which is a brand name for melphalan.
- A study design filter to identify clinical trials was applied to the Embase search. The filter was not referenced but appeared to be from Ovid Expert Searches. (http://resourcecenter.ovid.com/site/resources/expert_search/healthexp.html#OvidFilters)
- RCT study design filters were not applied to PubMed and Cochrane Library searches.
- A range of conference proceedings were also searched. Websites and a list of terms for these resources were provided.

4.1.2 Inclusion criteria

The inclusion/exclusion criteria for the clinical efficacy and safety component of this systematic review are reproduced in Table 4.2. The reported objective and inclusion/exclusion criteria indicate that the systematic literature review was limited to patients with one of four categories of haematological malignancies, namely acute lymphoblastic leukaemia (ALL), acute myeloid leukaemia (AML), myelodysplastic syndrome (MDS), and multiple myeloma (MM). The comparators listed in this table did not match either those defined in the final scope,[8] or those reported in the company's definition of the decision problem, as given in Document B[1] and reproduced in Table 3.1 above.

Appendix D of the company submission (CS) reported details of 28 publications, of which 18 related to studies using treosulfan in the conditioning regimen and 10 related to studies of conditioning regimens in stem cell transplants based on the listed comparators.[17] Not all of these studies were included/described in Document B of the submission,[1] or considered in the feasibility assessment for

the completion of indirect and mixed treatment comparisons to provide comparative evidence for TREO/FLU versus other conditioning regimens (reported in Appendix L of the CS).[18]

ERG comment: The company were asked to provide clarification on which of the publications listed in Appendix D of the CS related to the studies included in the main body of the CS (Document B, Section B.2, clinical effectiveness).[11] The company were also asked to explain what criteria were used to determine which of the studies identified in the systematic literature review should be included in Document B.[11] The following response was provided[12]:

"Our submission document B included those trials sponsored by medac along with any publication from the studies."

The ERG notes that those treosulfan studies not included in Document B are briefly described in Appendix D[17] and do not include any further RCTs with relevant comparators.

Furthermore, the company were asked to confirm whether other haematological malignancies were excluded and, if yes, to provide the reason for this exclusion.[11] The following response was provided[12]:

"According to the last EBMT report in 2016, a total of 14,260 alloHSCTs were performed for malignant indications. From these, 11,423 were performed in patients with AML (n = 6281), MDS (n = 1878), ALL (n = 2651) and MM (n = 433). Therefore, these diseases are representative for alloHSCT indications (~79% of all alloHSCTs in Europe in 2016) and the literature search concentrated on these malignancies.[19]

In addition, we believe that the effects of treosulfan are unrelated to the underlying condition as was noted in the Position Paper from Prof. Suttorp et al. in Appendix L of our submission. This Position Paper is the "Consensus reasoned opinion by haematologists on the role of Treosulfan for patients with malignant diseases undergoing allogeneic stem cell transplantation". Within this Position Paper, six European key opinion leaders (KOLs) (including 2 UK KOLs) and confirmed their opinion that the "advantages of TREO exist independently from the underlying disease to be treated." In this position paper they explain how studies across adults with AML/MDS, AML, MDS, lymphoid malignancies, chronic myeloid leukaemia and multiple myeloma have demonstrated overall survival and NRM which was "mostly superior or at least as good as other RIC or MAC regimens" in their study of the literature."

The ERG does not consider that this response provides sufficient justification for not searching for/excluding trials conducted in patients with malignancies other than AML, ALL, MDS and MM; according to the numbers provided, approximately 21% of alloHSCTs would be in patients with conditions other than those listed. The statement that the effects of treosulfan are unrelated to the underlying condition appears to be supported by clinical opinion only and, furthermore, is inconsistent with the exclusion of three studies from the feasibility assessment for network meta-analysis because they had "mixed populations" (see Table 1, Appendix L).[18, 20-22]

Recommended methods were used for initial inclusion screening (titles and abstracts): two reviewers independently assessed studies for inclusion and any disagreements were resolved through consultation with a senior reviewer. The company was asked to clarify whether full papers were also independently screened by two reviewers, and they confirmed that *"1st pass inclusion screening (titles and abstracts), data extraction and full-text review were completed independently by two reviewers"*.[12]

Table 4.2: Eligibility criteria for the systematic review of clinical effectiveness

Domain	Inclusion criteria	Exclusion criteria

Patient population	Adults, children and young people with malignant disease that is in **remission** before allogenic haematopoietic stem cell transplantation, including:
	• **Acute myeloid leukaemia (AML)**
	  • acute myeloblastic leukaemia
	  • acute promyelocytic leukaemia (APL/APML)
	  • acute myelocytic leukaemia
	  • acute myelomonocytic leukaemia
	  • acute myelomonoblastic leukaemia
	  • acute monoblastic/monocytic leukaemia
	  • acute/pure erythroid leukaemia
	  • acute erythroblastic leukaemia
	  • erythroleukaemia
	  • acute megakaryoblastic leukaemia
	  • acute basophilic leukaemia
	• **Acute lymphoblastic leukaemia (ALL)**
	  • B-lymphoblastic leukaemia (B-ALL)
	  • T-lymphoblastic leukaemia (T-ALL)
	  • NK cell lymphoblastic leukaemia
	• **Myelodysplastic syndrome (MDS)**
	• **Multiple myeloma (MM)**
	• **Acute leukaemia's of ambiguous lineage**
	  • acute undifferentiated leukaemia
	  • mixed phenotype acute leukaemia (MPAL)
	  • MPAL, B/myeloid, not otherwise specified (not otherwise specified [NOS])
		• Relapse/refractory/drug resistant disease
		• All types of lymphomas
		• Chronic lymphoblastic leukaemia (CLL)
		• Chronic myeloid leukaemia (CML)
		• Myelodysplastic/myeloproliferative neoplasms (MDS/MPN)
		• Myelofibrosis
		• Acute panmyelosis with myelofibrosis
		• Aplastic anaemia
		• Burkitt's Leukaemia

Domain	Inclusion criteria	Exclusion criteria
	• MPAL, T/myeloid, NOS	
Intervention	Treosulfan with fludarabine	-
Comparators	Conditioning therapy (high-dose or reduced intensity) with one of the following: • Cyclophosphamide and total body irradiation (TBI) • Cyclophosphamide and busulfan • Busulfan with fludarabine • Established clinical management without treosulfan with fludarabine • Standard of care • Melphalan • Thiotepa (THIO) • Anti-thymocyte globulin (ATG) • TBI + any other drug (even from the excluded list)	Regimens other than those in the include list, e.g.: • Bortezomib • Etoposide (VP16) • Teniposide (VM26) • Fractionated TBI • Clofarabine • Cytosine arabinoside • Cytarabine • Cladribine • Amsacrine • Methylchloride hexamethylene urea nitrate • Radionuclide-labelled antibodies e.g. ^{131}I-labeled anti-CD45
Outcomes	• Overall survival (OS) • Relative OS • Progression-free survival (PFS) • Disease-free survival (DFS) • Event-free survival (EFS) • Non-relapse mortality (NRM) • GvHD-free survival • Rates of relapse • Success of stem cell transplantation (engraftment)	• Genetic/polymorphism studies • Quality of life (flag)*

Domain	Inclusion criteria	Exclusion criteria
	• Adverse effects of treatment	
Study design	• RCTs • Non-RCTs (prospective or retrospective studies) – *for treosulfan studies only* • Case control studies	• Case report • Animal or *In-vitro* studies • Phase I trial • Phase I/II trial • Studies reporting the real-world effects
Date	2008-till date Congresses (2017 and 2018)	Publications earlier than 2008
Language	English	Any other language
Publication type	-	• Editorials • Commentaries • Letters • Protocol-only articles • Systematic reviews (flag)* • Reviews

Based on Table 1 of Appendix D of the CS[17]

* As reported in the CS. Unclear what "flag" refers to.

AML = acute myeloid leukaemia; ALL = acute lymphoid leukaemia; APL/APML = acute promyelocytic leukaemia; ATG = anti-thymocyte globulin; CLL = chronic lymphoblastic leukaemia; CML = chronic myeloid leukaemia; CS = company submission; DFS = disease-free survival; EFS = event-free survival; GvHD = Graft versus host disease; MDS = Myelodysplastic syndrome; MM = Multiple myeloma; MPAL = Mixed phenotype acute leukaemia MPN = Myeloproliferative neoplasms; NK = Natural killer; NOS = not otherwise specified; NRM = non-relapse mortality; OS = Overall survival; PFS = progression-free survival; RCT = randomised clinical trial; TBI = Total body irradiation; THIO = Thiotepa

4.1.3 Critique of data extraction

Appropriate measures to reduce the potential for error and bias in the data extraction process were reported in Appendix D of the CS[17]: *"Relevant data from the selected publications were extracted into predefined data extraction tables in Microsoft Word. All the data points were verified by an independent analyst."*

ERG comment: The ERG considers the methods used for data extraction to be sufficient.

4.1.4 Quality assessment

The company assessed the methodological quality of the identified RCT, MC-FludT.14/L Trial II, using a checklist designed to be used with multiple different study designs (RCTs, case–control studies, cohort studies, controlled before-and-after studies and interrupted time series) and recommended in Appendix F of the third edition of the methods for the development of NICE public health guidance [PMG4].[23] Document B of the CS[1] did not include any quality assessments for the three additional phase II prospective, non-randomised studies (MC-FludT.6/L, MC-FludT.7/AML and MC-FludT.8/MDS) listed and it was unclear whether any of the 21 quality assessments reported in Appendix D of the CS[17] were for publications relating to these studies.

No information was provided on the number of reviewers who assessed the quality of included studies.

ERG comment: Given that the only assessment of methodological quality included in Document B of the CS was for a randomised controlled trial (RCT) and that this RCT was the only study used in the cost effectiveness modelling, the ERG considers that the use of a risk of bias tool specifically designed for RCTs (e.g. the Cochrane risk of bias tool) may have been more informative.[24] The company were asked to confirm whether any such risk of bias assessments had been undertaken and to provide the results.[11] In order to minimise the potential for bias and error, it is usually recommended that two reviewers are involved in the extraction of data and assessment of study quality.

The company's response to clarification questions stated that: *"A total of 21 clinical studies (9 RCTs, 12 observational — 6 retrospective and 6 prospective) was assessed for quality appraisal using National Institute for Health and Care Excellence (NICE) recommended questionnaires for quantitative intervention (https://www.nice.org.uk/process/pmg4/chapter/appendix-f-quality-appraisal-checklist-quantitative-intervention-studies). The quality assessment was initially completed by one analyst and reviewed by a senior reviewer".*[12]

The ERG notes that the risk of bias assessment reported in Document B,[1] which is different to that reported in the systematic review described in Appendix D of the CS,[17] does include all of the components of the Cochrane risk of bias tool.[24, 25] The ERG considers that the criteria listed were appropriate for assessing risk of bias in an RCT.

4.1.5 Evidence synthesis

No meta-analyses, no indirect treatment comparisons nor mixed treatment comparisons were conducted.

The systematic literature review, reported in Appendix D of the CS,[17] identified four RCTs (including the main trial MC-FludT.14/L Trial II) that were deemed relevant for inclusion in any potential analysis in that they included regimens of interest (TREO/FLU), busulfan and fludarabine (BU/FLU), and busulfan and cyclophosphamide (BU/Cy)) and reported data for the efficacy endpoints reported for the main trial (i.e. overall survival, event-free survival, non-relapse mortality, relapse rate or graft-versus-

host disease).[10, 26-28] Appendix L of the CS reported a feasibility assessment for completion of a network meta-analysis and for indirect or mixed treatment comparisons based on these four trials.[18] This assessment concluded that it was not possible to perform a network meta-analysis for any of the efficacy outcomes reported in the main phase III trial (MC-FludT.14/L Trial II). Although the feasibility assessment identified that indirect comparisons may be possible for TREO/FLU versus BU/Cy and BU/FLU (MAC) at two years for overall survival (OS), relapse rate (RR) and the incidence of graft-versus-host disease (GvHD), the CS concluded that: *"these outcomes are unlikely to provide sufficient, reliable and relevant comparative data for inclusion in the economic assessment of TREO/FLU as a conditioning regimen for patients undergoing HSCT as treatment for malignant disease"*.[1]

ERG comment: The ERG considers that, wherever possible, any analyses that may provide comparative evidence for TREO/FLU versus other conditioning regimens of interest listed in the NICE scope should be performed. The company were asked to perform those indirect comparisons identified by the feasibility assessment, and to provide the results of these analyses.[11] The following response was provided[12]:

"As discussed in the response to B.2.d, an indirect comparison although technically possible for some of the endpoints was not thought to be sufficiently informative for the submission. The economic model is based on the pivotal trial (MC-FludT.14/L Trial II) and utilised the clinical endpoints of both OS and EFS. As the indirect comparison cannot provide data on EFS it has not been performed."

The ERG considers that an analysis of OS may have been possible, but acknowledges the differences between the trial populations.

4.2 Critique of trials of the technology of interest, the company's analysis and interpretation (and any standard meta-analyses of these)

Section B.2 of the CS identified one phase III RCT (MC-FludT.14/L Trial II),[10] and three phase II prospective non-randomised studies (MC-FludT.6/L,[26] MC FludT.7/AML[27] and MC-FludT.8/MDS)[28] of TREO/FLU as relevant to the submission. The phase III trial (MC-FludT.14/L Trial II) provides the main basis of the submission and is the only study clinical effectiveness study to be used in the cost effectiveness modelling.[10] The three phase II studies did not provide any additional information, either in terms of additional outcomes assessed or longer-term follow-up, and the following sections are therefore based on the main phase III RCT (MC-FludT.14/L Trial II).[10]

Furthermore, Section B.2 of the CS identified an ongoing phase II study (MC-FludT.17/M) of TREO/FLU in children.[29] In addition, Section B 2.2.4 of the CS reports results from an analysis of registry data intended to provide further comparative data for TREO/FLU versus other conditioning regimens; these results were not used in the cost effectiveness modelling.[1] These two studies are summarised in Sections 4.2.1 and 4.2.9 of this report.

4.2.1 Details of the included treosulfan RCT

The only included RCT of treosulfan (MC-FludT.14/L Trial II) assessed the efficacy and safety of treosulfan in combination with fludarabine for conditioning therapy as part of allogeneic haematopoietic stem cell transplantation (alloHSCT) in adults with AML or MDS who were at increased risk for standard conditioning therapies (i.e. not eligible for standard MAC busulfan- or TBI-based regimens).[10] The intervention was treosulfan (10 g/m²/day) in combination with fludarabine (30 mg/m²/day) and the comparator was the RIC regimen busulfan (3.2 mg/kg/day) also in combination with

fludarabine (30 mg/m²/day). A summary of study methodology for MC-FludT.14/L Trial II is provided in Table 4.3.

ERG comment: The ERG notes that the evidence for treosulfan in combination with fludarabine for conditioning therapy as part of alloHSCT is based on a multi-centre, international RCT investigating patient-relevant outcomes, with follow-up to two years after transplantation.

Restricted population

It should be noted that the study population was restricted to adults with AML or MDS who were at increased risk for standard conditioning therapies (i.e. not eligible for standard high-dose busulfan- or TBI-based regimens). This population is substantially narrower than that defined by NICE in the final scope[8] and in the decision problem, specified in Table 1 of the CS[1] namely *"Adults, children and young people with malignant disease that is in remission before allogenic haematopoietic stem cell transplantation"*. It is also narrower than the therapeutic indications reported in the summary of product characteristics (SmPC)[13]: *"Treosulfan in combination with fludarabine is indicated as part of conditioning treatment prior to allogeneic haematopoietic stem cell transplantation (alloHSCT) in adult patients with malignant and non-malignant diseases, and in paediatric patients older than one month with malignant diseases."* The company were asked to clarify whether the intended population for this submission is as described in the scope or the more restricted population, which is consistent with the evidence that has been presented.[11] The following response was provided:

"We have presented in our submission the best available evidence for treosulfan and this is based on the pivotal phase III trial (MC-FludT.14/L Trial II). This trial, with 570 randomised patients, was the largest ever prospective randomised clinical trial (RCT) comparing two conditioning regimens.[10] This trial focusses on "patients who cannot tolerate such MAC regimens and would otherwise be treated with less efficacious RIC regimens i.e. patients aged ≥ 50 years at transplant and/or with a Haematopoietic Cell Transplantation-Comorbidity Index (HCT-CI) score >2 (indicating an intermediate to high risk of co-morbidities which influence NRM)."

However, we believe the evidence from this population is applicable to the broader population of "Adults, children and young people with malignant disease that is in remission before allogenic haematopoietic stem cell transplantation" in line with the final scope. In Appendix L we provided the "Haematologists' Position Paper" from Prof. Suttorp et al. This position paper includes the opinion of six European KOLs (including 2 UK KOLs) and states:

"The signatories of this consensus statement wish to share their opinions and strong belief that TREO-based conditioning should not only be reserved to treat defined malignancies such as AML or MDS in certain age groups. We consider the existing trial data as compelling to support the use of TREO in most patients with malignant diseases requiring treatment with myeloablative conditioning followed by alloHSCT."

Therefore, we feel justified in believing that the trial population can be broadened in line with the Final Scope."

The ERG notes that the assertion that the evidence from MC-FludT.14/L Trial II is applicable to the broader population, defined in the scope and the decision problem, appears to be supported by clinical opinion alone. The ERG remains very concerned that population included in MC-FludT.14/L Trial II is not representative of the full population defined in the scope, specifically, there is a lack of evidence about the effectiveness of treosulfan/fludarabine conditioning regimens in people who are able to

tolerate standard MAC regimens, in adults with malignancies other than AML or MDS, and in children. Furthermore, the nature of the restricted population in MC-FludT.14/L Trial II (adults who were at increased risk and therefore not eligible for standard MAC conditioning regimens), and hence the choice of comparator (RIC Bu/Flu) means that the relative effectiveness and cost effectiveness of treosulfan/fludarabine has not been evaluated against the range of comparator regimens that would be relevant for the full population defined in the scope.

Limited comparison

The included RCT, MC-FludT.14/L Trial II, only provides comparative efficacy data for treosulfan in combination with fludarabine versus one alternative conditioning therapy, a RIC regimen of busulfan in combination with fludarabine. The company were asked to confirm whether the analysis of registry data (described in section B.2.2.4 of the CS) represented the only data being submitted on the comparative effectiveness of treosulfan with fludarabine vs. conditioning regimens other than busulfan in combination with fludarabine.[11] The following response was provided:

"Our submission focusses on reduced intensity conditioning because this is where we have direct phase III clinical trial evidence and also where there is an unmet need for a large proportion of patients. Further evidence for the comparison of treosulfan with fludarabine against cyclophosphamide, high-dose busulfan with thiotepa and melphalan plus fludarabine is based on the registry analysis which medac commissioned. The European Society for Blood and Marrow Transplantation (EBMT) registry was analysed and the results were discussed in section B.2.2.2.4 of the submission (a PDF is now provided as requested in question D2).[15]

In addition, the US Center for International Blood and Marrow Transplant Research (CIBMTR) was also recently asked by medac to analyse their registry and compare it to the data from the MC-FludT.14/L Trial II. The CIBMTR report[16] provides a comparison of allogeneic transplantation outcomes for acute myeloid leukemia (AML) and myelodysplastic syndrome (MDS) with full or reduced intensity busulfan with fludarabine regimens and busulfan-cyclophosphamide regimen in the United States to the results of the MC-FludT.14/L Trial II.

Lastly we refer you to the Haematologists' Position Paper within Appendix L in which six European Key Opinion Leaders summarise that "TREO/FLU conditioning leads to better outcomes than busulfan-cyclophosphamide, as well as melphalan-based regimens."

The ERG notes that the EBMT registry study included only patients between the ages of 50 and 70 years (i.e. those who would be classified as at increased risk and therefore not eligible for standard high-intensity MAC conditioning regimens).[15] The study report indicates that comparator (registry) patients received either fludarabine plus melphalan (Flu/Mel) or busulfan plus cyclophosphamide (Bu/Cy). The scope includes Flu/Mel as a relevant RIC comparator, however, the nature or the Bu/Cy regimen received by the registry patients is unclear; the dose was not specified and the study report indicates that patients received either Bu/Cy alone (strict population) or Bu/Cy plus various other drugs (not Bu/Cy with thiotepa, as indicated in the company's response above.[15] It is unclear whether this study included registry patients who were eligible for and received standard high-intensity MAC conditioning regimens, however, if such patients were included then they would, by definition, not be comparable to patients included in MC-FludT.14/L Trial II.

Increased risk patients

The study report for CIBMTR registry study,[16] was provided along with the company's response to clarification questions, but was not included in Document B of the CS[1] or utilised in the cost effectiveness modelling. The inclusion criteria for the comparator (registry) patients in this study indicate that only those aged ≥50 years or aged 18 to 70 with HCT-CI score >2 (i.e. meeting the definition of increased risk patients, not eligible for standard high-intensity MAC conditioning regimens, used in MC-FludT.14/L Trial II) were included. However, the conditioning regimens received by registry patients included in the study were: RIC Bu/Flu; RIC Bu/Flu plus ATG; MAC Bu/Flu; MAC Bu/Flu + anti-thymocyte globulin (ATG); MAC Bu/Cy. The ERG therefore questions whether MC-FludT.14/L Trial II used any additional criteria to define increased risk patients, not eligible for standard high-intensity MAC conditioning regimens.

Differences in inclusion criteria

As can be seen in Table 4.3, different exclusion criteria were applied in France to those used in the remaining four countries and study participants receiving matched unrelated donor (MUD) in France were given a different ATG preparation and regimen to that used in the remaining four countries. The company was asked to clarify whether any differences in the study results were observed (for any outcome measure) between geographic locations and to provide the results for any subgroup analyses (by country) conducted.[11] The following response was provided:

"Furthermore, only 14 patients were recruited from France into this study (~ 2.5% of all randomised patients). This low number could not compromise the overall study results and a subgroup analysis for such a small patient number would not be useful.

A country-specific sub-analysis of treatment results was not performed."

The ERG agrees with this response.

Representativeness of UK population

Regarding the extent to which the included RCT is representative of the UK population within the defined subgroup of people with malignant disease that is in remission before allogenic haematopoietic stem cell transplantation (i.e. adults with AML or MDS who were at increased risk for standard conditioning therapies and are therefore not eligible for standard high dose busulfan- or TBI-based regimens), the ERG notes that this trial does not include any UK patients. Section B.2.3.3 of the CS states that: *"The trial setting was comparable to standard NHS practice in the UK."*[1] However, the reference cited in support of this statement is for a NHS website which provides a general overview of stem cell and bone marrow transplants; i.e. the site includes general information about conditioning therapies but does not include any information about specific regimens or the frequency with which these are used in the UK NHS or about the types of patients receiving stem cell transplantation in the UK NHS (i.e. it does not provide sufficient information to support a meaningful comparison between the trial and standard NHS practice in the UK).[30] The company was asked: *"Compared to patients seen in practice in England and Wales, how similar does the company consider the patients in the key trial MC-FludT.14/L (conducted in France, Germany, Hungary, Italy and Poland) to be? Have any clinical experts commented on this issue? If so, please provide relevant documents".*[11] The following response was provided[12]:

"The general practice of alloHSCT is not different in England and Wales versus other major European countries (including France, Germany, Hungary, Italy and Poland). This is also demonstrated in the

most recent EBMT report which shows that 52 UK teams performed 4,316 alloHSCTs in 2016.[19] Since these UK transplant sites are members of the EBMT, they also treat their patients according to the EBMT Guidelines which are provided in the EBMT Handbook, version 2019.[5]

According to the latest British Society of Blood and Marrow Transplantation (BSBMT) registry (2017),[4] the most common indications for an alloHSCT in the UK are acute myeloid leukaemia (AML; 36%), acute lymphoblastic leukaemia (ALL; 16%) and myelodysplastic syndrome (MDS) or myelodysplastic/ myeloproliferative neoplasms (MDS/MPN; 13%). These indications for alloHSCT in the UK are comparable to those in adults and children undergoing HSCT in Europe (AML 38%; ALL 16%; and MDS or MDS/MPN 15%).[19] Further details were provided in section B.1.3.1. (Common haematological indications for HSCT in adults and children in the UK and section) B.1.3.1.3 (Patients requiring HSCT in the UK).

Within our submission in Appendix L we provided the "Haematologists' Position Paper" from Prof. Suttorp et al. As mentioned earlier, this Position Paper is the "Consensus reasoned opinion by haematologists on the role of Treosulfan for patients with malignant diseases undergoing allogeneic stem cell transplantation". Contributing experts for this document included also two transplant experts from the UK (Robert Wynn, Amit Patel). In their overall conclusion, these experts have stated that treosulfan-based conditioning should be made available for all patients with malignant diseases where an alloHSCT is indicated."

The ERG agrees that, in general, alloHSCT practice is likely to be similar in England and Wales to that in other major European countries. However, the ERG remains unclear as to how similar the definition of patients at increased risk for standard conditioning therapies (i.e. not eligible for standard high-intensity MAC) is to any such definition that would be generally applied in practice. Information from the reports of two registry studies (described above) indicates possible inconsistencies in this definition or its application.

Table 4.3: Summary of study methodology for the included RCT

	MC-FludT.14/L Trial II
Location	France, Germany, Hungary, Italy and Poland
Trial design	Randomised, parallel-group, open label, multicentre, international, group-sequential phase III non-inferiority trial

Inclusion criteria	Adult patients (age 18 to 70 years) at increased risk for standard conditioning therapies and who were, therefore, considered to be ineligible for registered standard high-dose regimens based on busulfan or TBI:
	• Patients with AML according to the WHO (AML in complete remission at transplant, i.e. blast counts < 5% in bone marrow) or MDS according to the WHO (MDS with blast counts < 20% in bone marrow during disease history) indicated for allogeneic haematopoietic progenitor cell transplantation but considered to be at increased risk for standard conditioning therapies according to the following criteria[31]:
	  • patients aged ≥ 50 years at transplant and/or
	  • patients with a HCT-CI score >2[32]
	• Availability of a human leukocyte antigen (HLA)-identical sibling donor (MRD) or HLA-identical unrelated donor (MUD). Donor selection was based on molecular high-resolution typing (4 digits) of class II alleles of the DRB1 and DQB1 gene loci and molecular (at least) low-resolution typing (2 digits) of class I alleles (i.e. antigens) of the HLA- A, B, and C gene loci. In case no class I and class II completely identical donor (10 out of 10 gene loci) could be identified, one antigen disparity (class I) and/or one allele disparity (class II) between patient and donor was acceptable. Conversely, disparity of 2 antigens (irrespective of the involved gene loci) was not accepted. These definitions for the required degree of histocompatibility applied to the selection of related as well as unrelated donors.
	• Karnofsky Index ≥ 60%
	• Men capable of reproduction and women of childbearing potential had to be willing to consent to using a highly effective method of birth control such as condoms, implants, injectables, combined oral contra-ceptives, intrauterine devices, sexual abstinence or vasectomised partner while on treatment and for at least 6 months thereafter.
	• Written informed consent

Exclusion criteria	Applied to Germany, Hungary, Italy, and Poland: • Patients with acute promyelocytic leukaemia with t(15;17)(q22;q12) and in first complete remission (CR1)

	Applied to France only:
	• Patients with acute promyelocytic leukaemia with t(15;17)(q22;q12) and in CR1
	• Patients with cytogenetic favourable acute myeloid leukaemia ("low risk" AML) and in CR1, who did not present unfavourable clinical or disease features like secondary or therapy-related AML or insufficient response to AML induction therapy
	• MDS patients with IPSS-R "very low risk" or "low risk" at trial entry, who did not present unfavourable clinical features during disease history like refractory severe thrombocytopenia with severe bleeding complications, life-threatening infectious complications due to severe neutropenia and/or very high red blood cell transfusion requirement and related complications.
	
	Applied to all countries:
	• Patients considered contra-indicated for allogeneic HSCT due to severe concomitant illness (within 3 weeks prior to scheduled Day -6):
	  • Patients with severe renal impairment like patients on dialysis or prior renal transplantation or S-creatinine > 3.0 x upper limit of normal (ULN) or calculated creatinine clearance < 60 ml/min
	  • Patients with severe pulmonary impairment, DLCOSB (Hb-adjusted)/or forced expiratory volume 1 second (FEV_1) < 50% or severe dyspnoea at rest or requiring oxygen supply
	  • Patients with severe cardiac impairment diagnosed by electrocardiogram (ECG) and left ventricular ejection fraction (LVEF) < 40%
	  • Patients with severe hepatic impairment indicated by hyperbilirubinaemia > 3 x ULN or ALT/AST > 5 x ULN
	• Active malignant involvement of the central nervous system (CNS)
	• HIV-positivity, active non-controlled infectious disease under treatment (no decrease of C-reactive protein [CRP] or procalcitonin [PCT]) including active viral liver infection
	• Previous allogeneic HSCT
	• Pleural effusion or ascites > 1.0 l
	• Pregnancy or lactation
	• Known hypersensitivity to treosulfan, busulfan and/or related ingredients
	• Participation in another experimental drug trial within 4 weeks prior to Day –6 of the protocol
	• Non-cooperative behaviour or non-compliance

	MC-FludT.14/L Trial II
	• Psychiatric diseases or conditions that might compromise the ability to give informed consent
Setting and locations where data were collected	The clinical trial was performed at 31 sites specialised in conducting alloHSCTs: 2 sites in France, 20 sites in Germany, 6 sites in Italy, 2 sites in Poland, and 1 site in Hungary. Patients were usually hospitalised from 1 week before start of conditioning up to successful engraftment of donor stem cells (an average for a total of 4 weeks).
Intervention(s)	Treosulfan i.v. (10 g/m²/day [d-4 to d-2]) + fludarabine i.v. (30 mg/m²/day [d -6 to d -2]) prior to alloHSCT
Comparator(s)	Dose-reduced busulfan plus fludarabine: Busulfan i.v. (3.2 mg/kg/day [d-4 to d-3]) + fludarabine i.v. (30 mg/m²/day [d-6 to d-2]) prior to alloHSCT
Permitted and disallowed concomitant medication	• The patients were not allowed to participate in another experimental drug trial within 4 weeks prior to Day –6 of the protocol. • Due to the complexity of conditioning treatment including cytotoxic therapy, pre- and post-transplant immunosuppression and other prophylactic treatments (to prevent infections, liver, renal or CNS toxicity), relevant concomitant treatments were standardised and declared mandatory in both trial arms. • In France, a different ATG preparation and regimen was used in MUD transplantation. This regimen is registered and was considered equivalent to the regimen used in the other countries. Other treatments, which were not standardised were to be conducted according to the centre-specific policy. • Relevant concomitant drug treatment given in the trial period between Day -6 and Day +28, i.e., conditioning treatments, prophylactic medication for HSOS, prophylactic medication for mucositis and growth factors (such as G-CSF, GM-CSF, rHU-KGF), were to be recorded on the CRF. • In addition, concomitant medication for prophylactic and/or therapeutic GvHD treatment was documented explicitly between Day -1 and Day +100 on the CRF. • Disease-specific interventions, which might have an impact on the primary trial objective (e.g., prophylactic or pre-emptive donor lymphocyte infusion (DLI), prophylactic/pre-emptive cytotoxic chemotherapy or radiotherapy after transplantation, but in the absence of relapse/disease progression), were not allowed.
Primary outcomes (including scoring methods and timings of assessments)	Event-free survival (EFS) 2 years after transplantation; EFS was measured from time of end of HSCT (= Day 0) to time of event. Events were defined as relapse of disease, graft failure or death (whatever occurred first).
Other outcomes used in the economic model/specified in the scope	• Comparative evaluation of incidence of CTC grade III/IV mucositis between Day -6 and Day +28

	MC-FludT.14/L Trial II
	• Comparative evaluation of overall survival (OS) and cumulative incidence of relapse (RI), non-relapse mortality (NRM) and transplantation-related mortality (TRM) within 2 years after transplantation. • Comparative evaluation of cumulative incidence of acute and chronic GvHD within 2 years after transplantation. • Comparative evaluation of incidence of other CTC grade III/IV adverse events between Day -6 and Day +28.
Pre-planned subgroups	Several subgroup analyses were planned including Cox proportional hazards regression model and Kaplan Meier analyses by donor type, risk group, combination of donor type and risk group, disease (AML vs MDS), age group (< 50 years vs ≥50 years), and HCT-CI Score (≤2 vs > 2)

Based on Tables 8 and 14, and section B.2.3 of the CS[1]
alloHSCT = allogeneic haematopoietic stem cell transplantation; ALT = Alanine aminotransferase; AML = acute myeloid leukaemia; AST = Aspartate aminotransferase; CNS = Central nervous system; CR1 = first complete remission; CRF = case report form; CRP = C-reactive protein; CS = company submission; CTC = Common Terminology Criteria; DLCOSB = single breath diffusing capacity for carbon monoxide; DLI = donor lymphocyte infusion; ECG = electrocardiogram; EFS = event-free survival; FEV_1 = forced expiratory volume 1 second; G-CSF = granulocyte colony-stimulating factor; GM-CSF = granulocyte-macrophage colony-stimulating factor; GvHD = graft versus host disease; HCT-CI = haematopoietic cell transplantation-comorbidity index; HIV = human immunodeficiency virus; HLA = human leukocyte antigen; HSCT = Haematopoietic stem cell transplantation; IPSS-R = International Prognostic Scoring System-Revised; MDS = myelodysplastic syndrome; LVEF = left ventricular ejection fraction; MRD = matched related donor; MUD = matched unrelated donor; NRM = non-relapse mortality; OS = Overall survival; PCT = procalcitonin; RCT = randomised clinical trial; rHU-KGF = recombinant human keratinocyte growth factor; TBI = total body irradiation; TRM = transplantation-related mortality; ULN = upper limit of normal; WHO = World Health Organization

4.2.2 Statistical analysis of the included treosulfan RCT

Details of the design and statistical analysis methods of the MC-FludT.14/L Trial II are provided in Table 4.4.

Table 4.4: Summary of statistical analyses

Trial number (acronym)	MC-FludT.14/L Trial II
Hypothesis objective	The primary objective of this randomised phase III trial was to demonstrate, as a minimum, non-inferiority of treosulfan as an alternative conditioning agent to busulfan with respect to EFS.
Statistical analysis	The non-inferiority margin on the hazard ratio scale was pre-specified as 1.3. If significant non-inferiority within the Per Protocol Set (PPS) could be shown, a sequential testing approach was to be applied starting with testing the non-inferiority within the Full Analysis Set (FAS). In case of statistical significance, superiority within the FAS with respect to the primary trial endpoint was to be tested based on the 'Points to Consider on Switching between Superiority and Non-inferiority (CPMP/EWP/482/99)'. For confirmatory analysis of non-inferiority of treosulfan-based conditioning a Cox proportional hazards regression model (stratified by centre and risk group) with donor type (MUD vs MRD) and treatment as factors was applied for EFS. These factors are exactly those factors used within the randomisation procedure. The analysis was based on the patients available after the second interim analysis. Beyond this confirmatory aim of the trial, exploratory data analyses were conducted. They consisted of a Cox proportional hazards regression model and Kaplan-Meier analyses by donor type, risk group, combination of donor type and risk group, disease (AML vs MDS), age group (< 50 years vs ≥ 50 years), and HCT-CI Score (≤2 vs > 2) for all patients included in the trial. To simplify descriptive comparisons, the subgroup analyses were graphically summarised by means of forest plots showing event-free survival by donor type, risk group, combination of donor type and risk group, disease (AML and MDS), age group (< 50 years vs ≥ 50 years), HCT-CI Score (≤2 vs >2), remission status in AML (CR1 vs > CR1), disease status at trial entry in MDS (untreated vs treated), risk group within AML patients, and risk group within MDS patients. For each subgroup they include the associated sample size, the number of events, the 24-month Kaplan-Meier estimates, the hazard ratio (HR) showing the risk of events with treosulfan compared to busulfan and the associated CI derived by means of Cox proportional hazards model with treatment as only factor. Sensitivity analyses were also performed to evaluate the robustness of the primary analysis. For the secondary endpoints, data for the FAS and PPS were analysed. For the endpoints overall survival, transplantation-related mortality, GvHD-free and relapse/progression-free survival, and chronic GvHD-free and relapse/progression-free survival and the exploratory endpoints time to deterioration of Karnofsky Performance Score (KPS) by at least 20 points and deterioration of KPS to less than 60 points, Kaplan-Meier estimates were calculated. A Cox proportional hazards regression model stratified by centre and risk group with donor type (MUD vs MRD) and treatment as factors was fitted. The adjusted estimate of treatment effect from the Cox proportional hazards model was expressed as a hazard ratio (treosulfan vs busulfan),

	together with the associated 95% CIs. In addition, the two-sided p-values based on Wald-test was conducted. For the endpoints relapse/progression, and non-relapse mortality, the probability over time was estimated by cumulative incidence rates. The test of Gray was applied to compare treatment arms. For engraftment, the conditional cumulative incidence was estimated using conditional probability functions. The two-sided Pepe-Mori test was used to compare treatment arms. Potential effects of covariates on the secondary endpoints were studied through subgroup analyses. The same subgroups and analysis techniques described above for event-free survival were considered. However, the non-inferiority testing and superiority testing was omitted because the trial was designed to test non-inferiority with respect to event-free survival only. The trial was planned as a group-sequential trial with 3 interim analyses. The first formal interim analysis was planned to be performed after 45 events or 220 patients to allow for a broad review of the benefits and risks of the dose reduction and change of the treatment regimen implemented with amendment 3 of the trial. Stopping early due to proof of efficacy or futility was unlikely within this interim analysis due to the low information fraction within this analysis. Further interim analyses were planned after 137 and 239 events occurred, or after 460 and 700 patients were randomised. The final analysis was planned after 481 events or inclusion of 930 patients at the latest. When reviewing the results of the second interim analysis, the DMC recommended to stop recruitment for this trial since the primary objective – the proof of non-inferiority of treosulfan compared to busulfan – had already been achieved. The Clinical Study Report (CSR) describes the results of the final analysis of all 570 patients enrolled in the trial and of these, 551 patients qualified for the FAS.
Sample size, power calculation	The sample size required within the scope of amendment 03 was calculated based on the hypothesis system described in the CSR (Section 9.7.1.1) applying an experiment-wise one-sided type-I-error significance $\alpha = 2.5\%$. Only those patients to be enrolled after implementation of the amendment 03 were subjected to confirmatory analysis, i.e. the 330 patients already recruited in MC-FludT.14/L Trial II prior to trial re-activation with amendment 03 were excluded. The rationale for this is twofold: All patients previously randomised to the treatment regimen with 14 g/m^2 treosulfan do not provide information for safety and efficacy of the newly developed regimen with 10 g/m^2. In addition, all patients previously randomised to busulfan may not be representative for future randomised patients due to potential selection- and performance-bias. Exclusion of any kind of bias is even more important in non-inferiority design settings. Sample size estimation assumed under the alternative hypothesis that treosulfan-based conditioning is equally effective to the comparator (i.e.: HR = 1). The power of the trial was 80%, so that the sponsor´s risk of erroneously overlooking the non-inferiority was 20%. Interim analyses for futility looks were incorporated in the revised protocol allowing for premature stop of the trial if it was unlikely to achieve the ultimate goal of the trial. In addition, the trial was to be stopped early if non-inferiority of treosulfan-based conditioning was clearly established. This resulted in a group-sequential approach with at most 3 confirmatory interim analyses and

	one final analysis, each with different stopping criteria (boundaries) for futility and efficacy. The resulting inflation of the overall Type I and Type II error probabilities was taken into account. The most conservative approach of interim efficacy monitoring by means of an O'Brien-Fleming type stopping boundary was applied. Based on these general conditions above, a commitment to at most 481 events within the FAS was given in the trial protocol. Since the power of any time-to-event trial is determined by the number of events rather than the number of patients, a range of sample sizes met the objectives of this trial. Naturally, the more patients being followed, the sooner the desired number of events is observed. Assuming, a 12-month EFS-rate of 68.5% with busulfan-based conditioning RIC (based on the results of first confirmatory interim analysis of MC-FludT.14/L Trial II prior to amendment 03) and anticipating (based on the accrual experience prior to amendment 03 and the already existent infrastructure of the trial) recruitment of 10 patients per month within the first 6 months, 15 patients per month thereafter until 24 months after re-start of the trial and 25 patients per month thereafter, the required number of events was expected to be reached with at most 930 patients. The maximum expected trial duration (accrual plus follow-up) to reach the required number of events was about 64 months after randomisation of the first patient with amendment 03 in force. The expected duration was 40 months under the null hypothesis and 58 months under the alternative hypothesis of non-inferiority. Assuming that roughly 3% of patients have to be excluded from FAS, at most 960 patients were to be enrolled in this trial.
Interim analysis	In order to stop the trial as soon as the question of non-inferiority could be answered, a group-sequential approach was implemented consisting of 3 confirmatory interim analyses. Interim analyses were conducted to allow for early stopping of the trial for significant non-inferiority as well as futility. In particular, the first interim analysis was performed with 220 randomised patients qualifying for FAS to investigate the effect of dose reduction on duration of neutropenia and TRM until Day +100. The second interim analysis was scheduled with 137 events or latest after randomisation of 460 patients qualifying for FAS. The third interim analysis was scheduled with 239 events or 700 patients in FAS. After the second interim analysis, the DMC recommended stopping recruitment into the trial since the primary objective of the trial, the proof of non-inferiority of treosulfan compared to busulfan had been achieved. Medac followed this recommendation and stopped recruitment after a total of 570 patients had been included. A confirmatory analysis based on 476 patients included in the second interim analysis was completed and presented in a previous clinical trial report. The final analysis of all 570 patients included in the trial, was performed after all patients had been followed-up for at least one year and when the post-surveillance documentation had been performed for patients who finished the trial alive after 2 years.
Data management,	Patients had to be permanently removed from the trial if they withdrew their consent. Only 5 of 570 patients withdrew their consent. Four of these patients

patient withdrawals	are included in the FAS. The other patient withdrew his consent prior to the start of treatment. If there was a permanent removal of a patient from the trial, all efforts were to be made to perform all assessments scheduled for the end-of-trial visit. The reason for withdrawal was documented on the CRF.
Based on CSR[26] AML = acute myeloid leukaemia; CI = confidence interval; CR1 = first complete remission CSR = clinical study report; DMC = data monitoring committee; EFS = event-free survival; FAS = full analysis set; GvHD = graft-versus-host disease; HCT-CI = haematopoietic cell transplantation co-morbidity index; HR = hazard ratio; KPS = Karnofsky Performance Score; MDS = myelodysplastic syndrome; MRD = matched related donor; MUD = matched unrelated donor; PPS = per protocol set; RIC = reduced intensity conditioning; TRM = transplantation-related mortality	

The analysis datasets were as follows:

- Safety which included all randomised patients treated at least once with trial medication;
- Full analysis set (FAS) which included all randomised patients from the safety set with at least one efficacy outcome measured after baseline.
- Per protocol set (PPS) which included all patients from the FAS who satisfied the following criteria:
    - Fulfilled all trial inclusion criteria and none of the exclusion criteria
    - Correct allocation to treatment group
    - Compliance with the administration of trial medication (a deviation of no more or less than 20% between the amount actually applied and the amount specified in the protocol)
    - Administration of ATG in case of MUD unless medical reasons for a deviation had been documented
    - Lack of any concomitant prophylactic/adjuvant DLI or cytotoxic therapy/radiotherapy after transplantation but in the absence of relapse/disease progression.

In all datasets, patients were analysed according to their randomised treatment group.

The trial was a non-inferiority trial designed to evaluate whether the treosulfan/fludarabine regimen was non-inferior to busulfan/fludarabine in terms of EFS at 24 months after alloHSCT. The trial design and analysis methods are appropriate for a non-inferiority trial with the main analysis of non-inferiority using the PP population followed by testing using the FAS. Confirmatory analyses of non-inferiority for EFS using Cox proportional hazards models stratified by centre and risk group and adjusting for donor type (MUD vs. MRD) were also performed. Although the PPS result is the primary one for demonstrating non-inferiority the FAS results were presented in the CS as non-inferiority of EFS was demonstrated in both populations.

Subgroup analyses were pre-specified for EFS and also used a Cox- proportional hazards model. The subgroups were donor type, risk group, the combination of donor type and risk group, disease (AML/MDS), age group (<50/ ≥50) and HCT-CI score (≤2/ >2).

Other trial outcomes were analysed using appropriate statistical methods, i.e. the ERG does not have any concerns about the design and analysis of the trial.

4.2.3 Participant characteristics for the included treosulfan studies

Table 4.5 shows the baseline characteristics of the participants in MC-FludT.14/L Trial II. The trial included a total of 570 participants. The majority of participants, 387/570 (67.9%), were recruited in Germany; 77/570 (13.5%) were recruited in Poland, 57/570 (10.0%) were recruited in Italy, 35/570 (6.1% were recruited in Hungary and 14 (2.5%) were recruited in France.[10]

Most study participants were 50 years of age or older, 523/551 (94.0% in the treosulfan group and 95.8% in the busulfan group), and the mean age of participants, across groups, was 59.6 ± 6.3 years. Across both treatment groups, the majority of patients (60.8%) were male. Demographic characteristics were similar in the two treatment groups.[1, 10]

In line with the specified inclusion criteria for the trial, all patients were at increased risk for the standard HSCT conditioning regimen; 94.9% were aged ≥50 years and 58.6% had an HCT-CI score >2.[1, 10]

Randomisation was stratified by donor type and risk group and thus, these factors were evenly distributed across the two treatment groups. Overall, more study participants had MUD transplantations, 421/551 (76.5%) than MRD 130/551 (23.5%), and more patients with AML, 352/551 (64.0%), than MDS, 199/551 (36.0%), were included in the trial. There were more patients with AML in the treosulfan group, 184/268 (68.6%), than the busulfan group, 168/283 (59.7%).[1, 10]

ERG comment: The ERG notes that the majority of trial participants were from Germany and that no information was provided about the ethnic distribution of study participants. Therefore, this aspect of the generalisibility of the trial population cannot be assessed.

4.2.4 Risk of bias assessment for the included treosulfan studies

The assessment of the methodological quality of MC-FludT.14/L Trial II (see Table 4.6), reported in Section B.2.5 of the CS,[1] used different criteria to those reported in the systematic review methods described in Appendix D of the CS.[17, 23] and no reference was provided for these criteria. The CS stated that: *"Overall, the risk of bias was considered to be low given that randomisation was robust and blinding was not possible for patients or treatment providers. There were no unexpected imbalances in withdrawals during the course of the study. Furthermore, the planned outcome measures were analysed and reported and most efficacy analyses used an intention to treat approach."*[1]

ERG comment: The ERG agrees with the company's assessment of risk of bias to the items on randomisation, allocation concealment, imbalances in drop-outs, selective outcome reporting and use of intention to treat (ITT) analysis, but notes that additional information, from the CSR (recorded in italics in Table 4.6), was required in order to assess randomisation and allocation concealment.[10, 26] The assessment reported in the CS confused the concept of allocation concealment, which concerns the concealment of the allocation sequence from those assigning participants to the intervention groups until the moment of assignment, with the concept of blinding (of participants, clinicians, or investigators) to treatment during the trial (after assignment).[1] The ERG considers that the possible effect of the imbalance in disease type between the treatment groups is unclear. While the ERG acknowledges that the nature of the treatments precludes blinding of participants and treating clinicians, it notes that the risk of bias associated with an open label type study remains high.

Table 4.5: Baseline characteristics of participants in MC-FludT.14/L Trial II

Baseline characteristic - Full Analysis Set (FAS)	Treosulfan (10 g/m²/day [d-4 to d-2])	Busulfan (3.2 mg/kg/day [d-4 to d-3])
Patients analysed, N= 551	**268**	**283**
Age, mean age, years (SD)	59.3 (6.5)	59.9 (6.0)
<50 years, n (%)	16 (6.0)	12 (4.2)
≥50 years, n (%)	252 (94.0)	271 (95.8)
Male, n (%)	162 (60.4)	173 (61.1)
Female, n (%)	106 (39.6)	110 (38.9)
Indications for alloHSCT		
AML, n (%)	184 (68.6)	168 (59.4)
MDS, n (%)	84 (31.3)	115 (40.6)
Disease status – AML, n (%)		
First complete remission (CR1)	159 (86.4)	144 (85.7)
>CR1	25(13.6)	24 (14.3)
Disease status – MDS, n (%)		
Untreated	42 (50.0)	47 (40.9)
Treated	42 (50.0)	68 (59.1)
Disease duration, mean (SD)		
AML (months)	8.16 (7.54)	7.57 (7.55)
MDS (months)	14.64 (22.81)	14.01 (18.09)
HCT-CI score at baseline, n (%)		
Patients with HCT-CI score >2	156 (58.2)	167 (59.0)
Donor type, n (%)		
MRD	62 (23.1)	68 (24.0)
MUD	206 (76.9)	215 (76.0)
MUD	206 (76.9)	215 (76.0)

Based on Table 15 of the CS[1]

alloHSCT = allogeneic haematopoietic stem cell transplantation; AML = acute myeloid leukaemia; CR1 = first complete remission; FAS = full analysis set; HCT-Cl: haematopoietic cell transplantation co-morbidity index; MDS = myelodysplastic syndrome; MRD = matched related donor; MUD = matched unrelated donor; SD = standard deviation

Table 4.6: Quality assessment for MC-FludT.14/L Trial II

Study question	How is the question addressed in the study?	Risk of bias, (High, Low, Unclear)	
		CS	ERG
Was randomisation carried out appropriately?	Randomisation was performed using a robust, validated block approach. The block length was unknown to transplant centres. To increase homogeneity between the 2 treatment arms, randomisation was stratified by cytogenetic and/or molecular risk group for AML, IPSS-R for MDS, as well as donor type. *Patients who met the enrolment criteria were centrally randomised by means of a computer-generated randomisation list. Randomisation was performed in a 1:1 ratio using a permuted block technique. The block length was unknown to the transplant centres.*[10]	Low	Low
Was the concealment of treatment allocation adequate?	Due to different treatment schedules with regard to the different infusion regimens as well as the additional, anticonvulsive treatment, which was mandatory in the reference group only, blinding of the trial medication was considered unfeasible within this orphan indication of high-risk patients. In addition, transplant centres usually constitute small units with limited staff only, so that a two-physician concept consisting of independent treating and rating physicians could not be implemented. The robust primary endpoint (EFS) of the trial, which is considered independent from the subjective view of the patient or the investigator, allows the conduct of the trial as an unblinded, but randomised open label trial. The sponsor and trial personnel were blinded to aggregated data evaluated during the trial. An independent CRO conducted the unblinded analyses needed for the DMC meetings and a statistician from this CRO presented the results at the DMC meetings. Internal statisticians and statistical programmers were unblended after finalisation of the Statistical Analysis Plan (SAP) and data release for the final confirmatory analysis on 31-May-2017. *Patients who met the enrolment criteria were centrally randomised by means of a computer-generated randomisation list. Randomisation was performed in a*	Low	Low

Study question	How is the question addressed in the study?	Risk of bias, (High, Low, Unclear)	
		CS	ERG
	1:1 ratio using a permuted block technique. The block length was unknown to the transplant centres.[10]		
Were the groups similar at the outset of the study in terms of prognostic factors, for example, severity of disease?	Demographic characteristics were similar in the 2 treatment groups. Within the FAS there were more male than female patients (60.8% male, 39.2% female). There were more patients with AML in the treosulfan group (68.6%) than the busulfan group (59.7%). However, disease status and risk group stratification for AML and MDS was comparable between the treatment groups. Stratified randomisation therefore resulted in well balanced treatment groups.	Low	Unclear
Were the care providers, participants and outcome assessors blind to treatment allocation? If any of these people were not blinded, what might be the likely impact on the risk of bias (for each outcome)?	Participants and healthcare providers could not be blinded as previously explained. However, sponsor and trial personnel were blinded to aggregated data evaluated during the trial. An independent CRO conducted the unblinded analyses needed for the DMC meetings and a statistician from this CRO presented the results at the DMC meetings.	Low	High
Were there any unexpected imbalances in drop-outs between groups? If so, were they explained or adjusted for?	No	Low	Low
Is there any evidence to suggest that the authors measured more outcomes than they reported?	No	Low	Low
Did the analysis include an intention-to-treat analysis? If so, was this appropriate and were appropriate methods used to account for missing data?	Yes, the patients within the FAS were analysed in their initial group of randomisation (intention to treat principle).	Low	Low
Did the authors of the study publication declare any conflicts of interest?	Study is unpublished.		
Based on Table 17 of the CS[1] and the CSR Medac, 25 Jan 2018 #234}			

Study question	How is the question addressed in the study?	Risk of bias, (High, Low, Unclear)	
		CS	ERG
AML = acute myeloid leukaemia; CRO = contract research organisation; CS = company submission; CSR = clinical study report; DMC = Data monitoring committee; EFS = event-free survival; ERG = Evidence Review Group; FAS = full analysis set; IPSS-R = International Prognostic Scoring System-Revised; MDS = myelodysplastic syndrome; SAP = statistical analysis plan			

4.2.5 Clinical effectiveness results for the included treosulfan studies

The efficacy results for the included RCT (MC-FludT.14/L Trial II) are shown in Table 4.7. This Table includes results for the primary outcome of the study, all secondary outcomes of the study that were listed in the final scope,[8] one outcome (NRM) which was not included in the scope but was included in the company's definition of the decision problem (see Section 3 of this report), and one additional, patient-relevant outcome (TRM) which was not included in the scope or the company's definition of the decision problem. All results are for the FAS population.[10]

The primary endpoint of MC-FludT.14/L Trial II was event-free survival within 24 months after alloHSCT. Events were defined as relapse of disease, graft failure, or death from any cause (whatever occurred first). Disease-specific definitions of relapse were provided in the CSR.[10]

Relapse definition acute myeloid leukaemia (AML):

- Morphologic relapse – Reappearance of leukaemic blasts in peripheral blood (PB) or ≥ 5% blasts in bone marrow (BM) after complete remission (CR), but not attributable to any other cause (e.g. bone marrow regeneration; if there were no blasts in PB and BM contained 5 to 20% blasts, a repeat BM evaluation performed at least a week later was necessary to distinguish relapse from BM regeneration); reappearance or development of cytologically proven extramedullary disease also indicated relapse.
- Cytogenetic relapse – Reappearance of cytogenetic abnormality, which had to be confirmed by a repeated diagnostic analysis prior to start of any therapeutic intervention **and** by an absolute decline in chimerism by ≥ 5%.
- Molecular relapse (only if a cytogenetic marker was not detectable) – Clinically relevant increase of molecular markers (proven by at least two documented evaluations) that had already been detected prior to patient inclusion and which required a therapeutic intervention (tapering/withdrawal of immunosuppression, DLI, cytotoxic or radio-therapeutic treatment).

Relapse/progression definition myelodysplastic syndrome (MDS):

- Increase in blasts to > 5% in BM or PB (if blasts > 5% at trial entry and patient experienced "marrow CR" after allogenic transplantation), but not attributable to other cause (e.g. bone marrow regeneration).
- Reappearance of cytogenetic abnormality (e.g. in case a MDS subtype without blasts in BM was included), which had to be confirmed by a repeated diagnostic analysis prior to start of any therapeutic intervention **and** by an absolute decline in chimerism by ≥ 5%.
- At least one of the following:
    - Decrement of ≥ 50% from maximum response level (after engraftment) in granulocytes or platelets in the absence of other conditions/reasons (e.g. antiviral or antibiotic or GvHD therapy) **and** an absolute decline of chimerism by ≥ 5%.
    - Reduction in haemoglobin concentration by ≥ 1.5 g/dL from maximum response level (after engraftment) or transfusion dependency **and** absence of other conditions/reasons (e.g. concomitant. antiviral or antibiotic or GvHD therapy or haemolysis due to AB0 incompatibility) **and** an absolute decline of chimerism by ≥ 5%.

Analyses, based on the FAS population, indicated that the treosulfan-based regimen was non-inferior to the busulfan-based regimen with regard to the primary endpoint EFS, and that the treatment

difference for EFS favoured treosulfan. The Kaplan-Meier curve for EFS in the FAS population is reproduced in Figure 4.1.

The overall survival, at 24 months after alloHSCT, was higher in the treosulfan treatment group compared with the busulfan treatment group. In addition, at the time of the post-surveillance evaluation, a higher proportion of patients in the treosulfan treatment group were alive compared with patients in the busulfan treatment group. The Kaplan-Meier curve for OS in the FAS population is reproduced in Figure 4.2.

Relapse/progression incidence was defined as the probability of having a relapse/progression. Patients were considered to have experienced an event when they relapsed/progressed. Death without relapse/progression and graft failures were competing risks. Patients alive with no history of relapse/progression were censored at the time of last clinical examination of disease status.[1] The cumulative incidence of relapse/progression (RI), in the FAS population, was comparable at 24 and 36 months after alloHSCT.

Non-relapse mortality was defined as the probability of dying without previous occurrence of a relapse or progression.[1] The cumulative incidence of NRM, in the FAS population, was lower in the treosulfan treatment group compared with the busulfan treatment group at 24 months and this difference persisted at 36 months.

Transplantation-related mortality was defined as all deaths occurring due to GvHD, cardiac toxicity, pulmonary toxicity, interstitial pneumonitis, haemorrhage, hepatic sinusoidal obstruction syndrome (HSOS), skin toxicity, Epstein-Barr virus (EBV) proliferative disease, renal failure, gastrointestinal toxicity, rejection/poor graft function, central nervous system (CNS) toxicity, multiple organ failure, infections (bacterial, viral, fungal, parasitic, unknown), or other HSCT-related causes.[1] For the FAS population, transplantation-related mortality was lower in the treosulfan treatment group compared with the busulfan treatment group at 24 months after alloHSCT. In addition, there was little change in TRM from 12 months to 24 months in treosulfan group (11.7% to 12.8%), whereas, in the busulfan treatment group, TRM increased from 16.2% to 24.1% from 12 months to 24 months.

ERG comment: The ERG notes that the key evidence presented in the clinical effectiveness section of the CS,[1] which was also used to inform cost effectiveness modelling, comprised a single RCT (MC-FludT.14/L Trial II) conducted in a population more restricted than that defined in the final scope and the company's definition of the decision problem (see Section 3 of this report).[1, 8] The company were, therefore, asked to confirm that they were not aware of any studies of the comparative effectiveness of treosulfan with fludarabine vs. other conditioning regimens for malignant disease that is in remission before allogenic haematopoietic stem cell transplantation that have been conducted in:

a. paediatric populations,
b. adults with haematological malignancies other than AML and MDS (those included in the key trial MC-FludT.14/L Trial II), and
c. adults who would be eligible for standard MAC (patients included in the key trial MC-FludT.14/L Trial II "were not eligible for a standard MAC busulfan- or TBI-based regimen").[11]

The company confirmed that *"no other prospective randomised studies with treosulfan-based conditioning versus other conditioning regimens are available."*[12]

The response also included brief summaries of five retrospective comparisons,[33-37] of which three were studies identified by the systematic literature review, reported in Appendix D of the CS.[17] None of these studies were utilised in the cost effectiveness modelling.

Table 4.7: Efficacy results of MC-FludT.14/L Trial II

	Treosulfan (10 g/m²/day) + Fludarabine (30 g/m²/day)	**Busulfan (3.2 mg/kg/day) + Fludarabine (30 g/m²/day)**
Number randomised	280	290
Number analysed (FAS: patients who received conditioning treatment and HSCT)	268	283
Median follow-up[a], months (range of those surviving)	29.7 (3.0 to 52.1)	29.4 (3.0 to 54.3)
Primary outcome – Event-free survival (EFS) within 24 months after alloHSCT		
Patients with event	97 (36.2%)	137 (48.4%)
Death[b]	35 (13.1%)	56 (19.8%)
Relapse/progression[b]	61 (22.8%)	72 (25.4%)
Primary graft failure[b]	1 (0.4%)	1 (0.4%)
Secondary graft failure[b]	0 (0.0%)	8 (2.8%)
Event-free survival at 12 months[c] [%] (95% CI)	70.0 (64.1 to 75.1)	60.8 (54.9 to 66.3)
Event-free survival at 24 months[c] [%] (95% CI)	65.7 (59.5 to 71.2)	51.2 (45.0 to 57.0)
Event-free survival at 36 months[c] [%] (95% CI)	[redacted]	[redacted]
Hazard ratio [HR][d] (95% CI)	0.64 (0.49 to 0.84)	
Secondary outcome – Overall survival (OS) within 24 months after alloHSCT		
Patients with event	81 (30.2%)	112 (39.6%)
Overall survival at 12 months[c] [%] (95% CI)	77.8 (72.3 to 82.3)	71.8 (66.1 to 76.7)
Overall survival at 24 months[c] [%] (95% CI)	72.7 (66.8 to 77.8)	60.2 (54.0 to 65.8)
Overall survival at 36 months[c] [%] (95% CI)	[redacted]	[redacted]
HR[d] (95% CI)	0.64 (0.48 to 0.87)	
Secondary outcome – Cumulative incidence of relapse/progression 24 months after HSCT		
Patients with event	61 (22.8%)	72 (25.4%)

	Treosulfan (10 g/m²/day) + Fludarabine (30 g/m²/day)	Busulfan (3.2 mg/kg/day) + Fludarabine (30 g/m²/day)
Patients without event (censored) or with competing event	207 (77.2%)	211 (74.6%)
Censored	171 (63.8%)	146 (51.6%)
Death[b]	35 (13.1%)	56 (19.8%)
Primary graft failure[b]	1 (0.4%)	1 (0.4%)
Secondary graft failure[b]	0 (0.0%)	8 (2.8%)
Cumulative incidence at 12 months [%] (95% CI)	19.1 (14.4 to 23.8)	21.7 (16.9 to 26.5)
Cumulative incidence at 24 months [%] (95% CI)	22.0 (16.9 to 27.1)	25.2 (20.0 to 30.3)
Cumulative incidence at 36 months [%] (95% CI)	[redacted]	[redacted]
HR[e] (95% CI)	0.82 (0.59 to 1.16)	
Secondary outcome - engraftment		
Primary graft failure[f]	1/268 (0.4%)	1/283 (0.4%)
Secondary graft failure[f]	0/263 (0.0%)	8/279 (2.9%)
Secondary outcome (not specified in scope) – Cumulative incidence of non-relapse mortality (NRM) 24 months after HSCT		
Patients with event	35 (13.1%)	56 (19.8%)
Patients without event (censored) or with competing event	233 (86.9%)	227 (80.2%)
Censored	171 (63.8%)	146 (51.6%)
Relapse/Progression[b]	61 (22.8%)	72 (25.4%)
Primary Graft Failure[b]	1 (0.4%)	1 (0.4%)
Secondary Graft Failure[b]	0 (0.0%)	8 (2.8%)
Cumulative incidence at 12 months [%] (95% CI)	10.5 (6.8 to 14.2)	14.3 (10.2 to 18.4)
Cumulative incidence at 24 months [%] (95% CI)	12.0 (8.0 to 15.9)	20.4 (15.5 to 25.2)
Cumulative incidence at 36 months [%] (95% CI)	[redacted]	[redacted]
HR[g] (95% CI)	0.63 (0.41 to 0.97)	

	Treosulfan (10 g/m²/day) + Fludarabine (30 g/m²/day)	Busulfan (3.2 mg/kg/day) + Fludarabine (30 g/m²/day)
Secondary outcome (not specified in scope) – Transplantation-related mortality (TRM)		
Patients with event	33 (12.3%)	58 (20.5%)
Patients without event	235 (87.7%)	225 (79.5%)
Transplantation-related mortality at 12 months[c] [%] (95% CI)	11.7 (8.3 to 16.3)	16.2 (12.2 to 21.3)
Transplantation-related mortality at 24 months[c] [%] (95% CI)	12.8 (9.2 to 17.7)	24.1 (19.1 to 30.2)
HR[d] (95% CI)	0.52 (0.34 to 0.82)	

Based on Tables 18, 19, 20, 21 and 23 of the CS[1]

[a] Based on reverse Kaplan-Meier estimates for overall survival; [b] Only if this event occurred first; [c] Based on Kaplan-Meier estimates; [d] Adjusted for donor type as factor, and risk group and centre as strata using Cox regression model; [e] Adjusted for donor type as factor and risk group as stratum using Fine and Gray model; [f] Rate of primary/secondary graft failure calculated as number of patients with graft failure by the number of patients at risk; [g] Adjusted for donor type as factor

alloHSCT = allogeneic haematopoietic stem cell transplantation; CI = confidence interval; CS = company submission; EFS = event-free survival; FAS = full analysis set; HR = hazard ratio; HSCT = Haematopoietic stem cell transplantation; NRM = non-relapse mortality; OS = Overall survival; TRM = transplantation-related mortality

Figure 4.1: Kaplan Meier estimates of EFS (FAS) MC-FludT.14/L Trial II

Figure 4 of the CS[1]
a Adjusted for donor type as factor, and risk group and centre as strata using Cox regression model; b For testing non-inferiority of treosulfan compared with busulfan; c For testing superiority of treosulfan compared with busulfan
CI = confidence interval; CS = company submission; FAS = full analysis set

Figure 4.2: Kaplan Meier estimates of OS (FAS) MC-FludT.14/L Trial II

Based on Figure 11.4.1.2.A of the CSR[10]
a Adjusted for donor type as factor, and risk group and centre as strata using Cox regression model; b For testing non-inferiority of treosulfan compared with busulfan; c For testing superiority of treosulfan compared with busulfan
CI = confidence interval; CSR = clinical study report; FAS = full analysis set

4.2.6 Subgroup analyses for the included treosulfan studies

The CS[1] states that subgroup analyses were implemented, in MC-FludT.14/L Trial II, in order to investigate the consistency of the study results; subgroups were specified in the statistical analysis plan (SAP).[38] The following subgroups were considered for EFS and OS: centre, risk group, donor type (MUD vs MRD), combination of donor type and risk group, disease (AML vs MDS), remission status at study entry (CR1 vs > CR1, for AML only), disease status at study entry (treated vs untreated, for MDS only), age group (<50 years vs ≥50 years), and HCT-CI Score (≤2 vs >2).[1]

The overall benefit of treosulfan, in terms of EFS and OS, was generally consistent across subgroup analyses.

[redacted].[1] The results of the subgroup analyses, for EFS and OS, in the FAS population are summarised in Figures 4.3. and 4.4.

Figure 4.3: Forest plot for EFS by prognostic factors with 24 month event rates (FAS) MC-FludT.14/L Trial II

Based on Figure 5 of the CS[1]
*Adjusted for donor type as factor, and risk group and centre as strata using Cox regression model
AML = acute myeloid leukaemia; CI = confidence interval; CS = company submission; EFS = event-free survival; FAS = full analysis set; HCT-CI = hematopoietic cell transplantation-comorbidity index; HR = hazard ratio; MDS = myelodysplastic syndrome; MRD = matched related donor; MUD = matched unrelated donor; N = number of participants; n: total number of events; RG = risk group

Figure 4.4: Forest plot for OS by prognostic factors with 24 month event rates (FAS) MC-FludT.14/L Trial II

Based on Figure 6 of the CS[1]
*Adjusted for donor type as factor, and risk group and centre as strata using Cox regression model
AML = acute myeloid leukaemia; CI = confidence interval; CS = company submission; FAS = full analysis set; HCT-CI = hematopoietic cell transplantation-comorbidity index; HR = hazard ratio; MDS = myelodysplastic syndrome; MRD = matched related donor; MUD = matched unrelated donor; N = number of participants; n: total number of events; OS = overall survival; RG = risk group

4.2.7 Health-related quality of life results for the included treosulfan studies

The included RCT (MC-FludT.14/L Trial II) did not assess any health-related quality of life (HRQoL) outcomes and no other studies reported HRQoL outcomes were included in the clinical effectiveness section of the CS.[1, 10]

4.2.8 Safety results for the included treosulfan studies

This section considers the information about adverse events (AEs) provided in the CS. All adverse events data were derived from MC-FludT.17/M. The safety analysis set (SAS) included all randomised patients who were treated at least once with trial medication; all patients were analysed within their group of actual treatment. Table 4.8 provides an overall summary of the frequency and severity (based

on Common Terminology Criteria for Adverse Events (CTCAE)) of adverse events; an AE was classified as treatment-related if the relationship was classified as "possibly related" or "related" by the Investigator.[10] Symptoms definitely caused by episodes of GvHD were documented separately.[1]

The frequency and severity of AEs was similar between the two treatment groups.

Table 4.8: Summary of treatment emergent adverse events MC FludT.14/L

Safety analysis set (SAS)	Busulfan (N=283)	Treosulfan (N=270)	Total (N=553)
Any AE, n (%)			
Patients with any AE	272 (96.1)	250 (92.6)	522 (94.4)
Patients with AEs of at least CTCAE Grade III	151 (53.4)	148 (54.8)	299 (54.1)
Drug-related AEs, n (%)			
Patients with any drug-related AEs	192 (67.8)	170 (63.0)	362 (65.5)
Patients with drug-related AEs of at least CTCAE Grade III	82 (29.0)	72 (26.7)	150 (27.8)
SAE, n (%)			
Patients with any SAE	20 (7.1)	23 (8.5)	43 (7.8)
Resulting in death	6 (2.1)	8 (3.0)	14 (2.5)
Life-threatening	8 (2.8)	13 (4.8)	21 (3.8)
Hospitalisation or prolongation of hospitalisation	9 (3.2)	8 (3.0)	17 (3.1)
Drug-related SAEs, n (%)			
Patients with any drug related SAE	9 (3.2%)	9 (3.3%)	18 (3.3%)
Maximum CTCAE grade of adverse events [n (%)]			
Patients with AEs of a maximum CTCAE grade I	46 (16.3%)	41 (15.2%)	87 (15.7%)
Patients with AEs of a maximum CTCAE grade II	75 (26.5%)	61 (22.6%)	136 (24.6%)
Patients with AEs of a maximum CTCAE grade III	134 (47.3%)	123 (45.6%)	257 (46.5%)
Patients with AEs of a maximum CTCAE grade IV	14 (4.9%)	18 (6.7%)	32 (5.8%)
Patients with AEs of a maximum CTCAE grade V	3 (1.1%)	7 (2.6%)	10 (1.8%)

Based on Table 30 of the CS[1]
AE = adverse event; CS = company submission; CTCAE = Common Terminology Criteria for Adverse Events; N = number of participants; n = number of patients in category; SAE = Serious adverse event; SAS = safety analysis set

Table 4.9 provides a more detailed breakdown of AEs, summarising the frequency of treatment-emergent adverse events (TEAE) TEAE of Grade III or higher by CTCAE System Organ Class and Term occurring in ≥5% of patients in either treatment group.[1] The overall rates of TEAEs of CTCAE Grade III or higher were similar in the two treatment groups; gastrointestinal disorders were more common in the busulfan group, and infestations and infections were more common in the treosulfan group.[1]

Drug-related TEAEs of Grade III or higher were reported by approximately a third of patients in both treatment groups (treosulfan, 26.7%; busulfan, 29.0%). Such events were reported by ≥5% of patients

in either treatment group for two TEAEs classes: investigations (treosulfan, 10.4%; busulfan, 9.5%) and gastrointestinal disorders (treosulfan, 11.3%; busulfan, 9.6%).[1]

Table 4.9: Frequency of CTCAE Grade III or above treatment emergent adverse events occurring in at least 5% of patients in either treatment group MC-FludT.14/L Trial II

CTCAE System Organ Class/Term	Busulfan (N=283)	Treosulfan (N=270)	Total (N=553)
Patients with any event	151 (53.4%)	148 (54.8%)	299 (54.1%)
Gastrointestinal disorders			
Any event	44 (15.5%)	33 (12.2%)	77 (13.9%)
Mucositis oral	21 (7.4%)	16 (5.9%)	37 (6.7%)
Nausea	17 (6.0%)	8 (3.0%)	25 (4.5%)
Investigations			
Any event	38 (13.4%)	39 (14.4%)	77 (13.9%)
Gamma-glutamyl transferase increased	25 (8.8%)	12 (4.4%)	37 (6.7%)
Alanine aminotransferase increased	9 (3.2%)	14 (5.2%)	23 (4.2%)
Blood and lymphatic system disorders			
Any event	31 (11.0%)	40 (14.8%)	71 (12.8%)
Febrile neutropenia	31 (11.0%)	40 (14.8%)	71 (12.8%)
Infections and infestations			
Any event	26 (9.2%)	41 (15.2%)	67 (12.1%)
Vascular disorders			
Any event	36 (12.7%)	27 (10.0%)	63 (11.4%)
Hypertension	27 (9.5%)	21 (7.8%)	48 (8.7%)
Metabolism and nutrition disorders			
Any event	16 (5.7%)	24 (8.9%)	40 (7.2%)

Based on Table 31 of the CS[1]
CS = company submission; CTCAE = Common Terminology Criteria for Adverse Events; N = number of participants

The incidence of GvHD was reported for the FAS population. GvHD occurring up to day 100 after HSCT was classified as acute and GvHD after day 100 was classified as chronic.[1] Incidence of acute and chronic GvHD are summarised in Tables 4.10 and 4.11, respectively. Although treosulfan appears to be associated with some numerical advantage compared to busulfan, for acute GvHD and extensive chronic GvHD, the difference was uncertain and did not reach statistical significance for any outcome. There was no statistically significant difference in the overall incidence of GvHD-related deaths between the two treatment groups (busulfan 7.4%, treosulfan 4.8%).[1]

Table 4.10: Frequency of acute GvHD (FAS) MC-FludT.14/L Trial II

	Treosulfan (N=268)	Busulfan (N=283)
Acute GvHD grade I-IV		
Patients with event	141 (52.6%)	162 (57.2%)
Patients without event (censored) or with competing event	127 (47.4%)	121 (42.8%)
Censored	101 (37.7%)	97 (34.3%)
Death[a]	11 (4.1%)	4 (1.4%)
Relapse/Progression[a]	14 (5.2%)	17 (6.0%)
Primary Graft Failure[a]	1 (0.4%)	1 (0.4%)
Secondary Graft Failure[a]	0 (0.0%)	2 (0.7%)
Cumulative incidence at 14 days [%] (95% CI)	10.5 (6.8 to 14.2)	14.1 (10.1 to 18.2)
Cumulative incidence at 28 days [%] (95% CI)	30.0 (24.5 to 35.5)	36.7 (31.1 to 42.4)
Cumulative incidence at 100 days [%] (95% CI)	52.8 (46.8 to 58.8)	57.2 (51.5 to 63.0)
HR (95% CI)	0.87 (0.69 to 1.08)	
Acute GvHD grade III-IV		
Patients with event	17 (6.3%)	23 (8.1%)
Patients without event (censored) or with competing event	251 (93.7%)	260 (91.9%)
Censored	214 (79.9%)	215 (76.0%)
Death[a]	11 (4.1%)	7 (2.5%)
Relapse/Progression[a]	25 (9.3%)	34 (12.0%)
Primary Graft Failure[a]	1 (0.4%)	1 (0.4%)
Secondary Graft Failure[a]	0 (0.0%)	3 (1.1%)
Cumulative incidence at 14 days [%] (95% CI)	0.7 (0.0 to 1.8)	0.7 (0.0 to 1.7)
Cumulative incidence at 28 days [%] (95% CI)	1.1 (0.0 to 2.4)	2.8 (0.9 to 4.8)
Cumulative incidence at 100 days [%] (95% CI)	6.4 (3.4 to 9.3)	8.1 (4.9 to 11.3)
HR (95% CI)	0.78 (0.42 to 1.45)	

Based on Table 33 of the CS[1]
[a] Only if this event occurred first
CI = confidence interval; CS = company submission; FAS = full analysis set; GvHD = graft-versus-host disease; HR = Hazard ration; N = number of participants

Table 4.11: Frequency of chronic GvHD (FAS) MC-FludT.14/L Trial II

	Treosulfan (N=268)	Busulfan (N=283)
Chronic GvHD		
Patients at risk[a]	229	232
Patients with event	138 (60.3%)	138 (59.5%)
Patients without event (censored) or with competing event	91 (39.7%)	94 (40.5%)

	Treosulfan (N=268)	Busulfan (N=283)
Censored	59 (25.8%)	58 (25.0%)
Death[b]	9 (3.9%)	9 (3.9%)
Relapse/Progression[b]	23 (10.0%)	22 (9.5%)
Primary Graft Failure[b]	0 (0.0%)	0 (0.0%)
Secondary Graft Failure[b]	0 (0.0%)	5 (2.2%)
Cumulative incidence at 6 months [%] (95% CI)	40.3 (34.0 to 46.7)	41.9 (35.5 to 48.2)
Cumulative incidence at 12 months [%] (95% CI)	54.8 (48.4 to 61.3)	55.1 (48.7 to 61.5)
Cumulative incidence at 24 months [%] (95% CI)	61.7 (55.1 to 68.3)	60.3 (53.8 to 66.7)
HR (95% CI)	1.00 (0.79 to 1.27)	
Extensive chronic GvHD		
Patients at risk[a]	229	232
Patients with event	45 (19.7%)	62 (26.7%)
Patients without event (censored) or with competing event	184 (80.3%)	170 (73.3%)
Censored	140 (61.1%)	110 (47.4%)
Death[b]	13 (5.7%)	24 (10.3%)
Relapse/Progression[b]	31 (13.5%)	31 (13.4%)
Primary Graft Failure[b]	0 (0.0%)	0 (0.0%)
Secondary Graft Failure[b]	0 (0.0%)	5 (2.2%)
Cumulative incidence at 6 months [%] (95% CI)	11.4 (7.3, 15.5)	13.8 (9.4, 18.2)
Cumulative incidence at 12 months [%] (95% CI)	16.7 (11.9 to 21.6)	20.0 (14.8 to 25.2)
Cumulative incidence at 24 months [%] (95% CI)	19.8 (14.5 to 25.1)	28.6 (22.5 to 34.7)
HR (95% CI)	0.71 (0.48 to 1.04)	

Based on Table 34 of the CS[1]
[a] Patients are at risk if they have survived 100 days after end of HSCT without relapse and graft failure.
[b] Only if this event occurred first
CI = confidence interval; CS = company submission; FAS = full analysis set; GvHD = graft-versus-host disease; HR = Hazard ration; N = number of participants

ERG comment: Overall, the ERG agrees with the company's conclusions that the safety data from Mc-FludT.14/L Trial II indicate that adverse event rates, incidence of acute and chronic GvHD, and GvHD-related deaths were similar in the treosulfan and busulfan groups.

The ERG notes that safety data were lacking for treosulfan used in the paediatric population with malignant disease that is in remission before allogenic haematopoietic stem cell transplantation.

In addition, data were lacking for the comparative safety of treosulfan with fludarabine vs. alternative conditioning regimens (other than RIC busulfan with fludarabine); no safety data were provided for comparator conditioning regimens other than RIC busulfan with fludarabine.

4.2.9 Supporting evidence from additional/ongoing studies

Section B.2.2.3 of the CS[1] described an ongoing phase II, prospective, single arm, open-label, multicentre, non-controlled study (MC-FludT.17/M) designed to assess safety and efficacy of treosulfan as part of a standardised fludarabine-containing conditioning therapy prior to alloHSCT in a paediatric population.[10] The study enrolled paediatric patients aged 28 days to <18 years with haematological malignancies (i.e. ALL, AML, MDS, or juvenile myelomonocytic leukaemias [JMML]), indicated for alloHSCT. A summary of study methodology for MC-FludT. 17/M is provided in Table 4.12.

Table 4.12: Summary of study methodology for the ongoing paediatric study (MC-FludT.17/M)

Location	Czech Republic, Germany, Italy, Poland, United Kingdom
Trial design	Phase II, prospective, single arm, open-label, multicentre, non-controlled study
Number of patients	N=70
Eligibility criteria for participants	Children up to 17 years with Haematological malignant disease i.e. ALL, AML, MDS or JMML, indicated for alloHSCT
Settings and locations where the data were collected	Specialist HSCT centres
Trial drugs	Treosulfan: i.v., BSA adapted: 10, 12 or 14 g/m²/day within 120 min to be administered prior to fludarabine; Fludarabine: i.v., 30 mg/m2/day on days from -7 to -3 prior to HSCT. With or without Thiotepa: i.v., 2 x 5 mg/kg/day on day -2 depending on investigator choice
Estimated study completion date	Last subject completes long-term follow-up in September 2019

Based on Table 10 of the CS[1]
ALL = acute lymphoblastic leukaemia; alloHSCT = allogeneic haematopoietic stem cell transplantation AML = acute myeloid leukaemia; CS = company submission; HSCT = Haematopoietic stem cell transplantation; i.v. = intravenous; JMML = juvenile myelomonocytic leukaemias; MDS = myelodysplastic syndrome; mg = milligram

The rate for freedom from transplant (treatment)-related mortality until 100 days after HSCT, the primary outcome for this study, was [redacted]% (90% CI: [redacted]%).[1]

The study is not yet completed as long-term (three years) follow-up is on-going for some patients. As of analysis of 12 March 2018, the key results relevant to the effectiveness outcomes specified in the scope for this appraisal[8] or included in the company's definition of the decision problem, were:

- Fourteen participants ([redacted]%) had experienced relapse / progression, graft failure or death. The Kaplan-Meier estimate of EFS at 12 months was [redacted]% (90% CI: [redacted]).[1]
- Sixty-three participants ([redacted]%) were alive. The Kaplan-Meier estimate of OS at 12 months after HSCT was [redacted]% (90% CI: [redacted]).[1]
- Eleven participants ([redacted]%) had experienced disease relapse/progression. The cumulative incidence of relapse / progression at 12 months was [redacted]% (90% CI: [redacted]).[1]
- Two participants ([redacted]%) had died without relapse. The cumulative incidence of NRM at 12 months was [redacted]% (90% CI: [redacted]).[1]

- No participants experienced a primary graft failure, and [redacted] subject ([redacted]%) experienced a secondary graft failure.[1]

The CS stated that: *"Overall, the reported safety and efficacy results of this Phase 2 alloHSCT trial demonstrated a positive benefit-risk for the treosulfan-based conditioning regimen used in the selected paediatric population and thus allowing extension of the use of treosulfan to this paediatric population."*[1]

ERG comment: The ERG does not consider that this ongoing, phase II study provides sufficient evidence to support the extension of the use of the treosulfan-based conditioning regimen to the paediatric population with haematological malignancies, because it does not provide any indication of comparative effectiveness vs. other alternative conditioning regimens.

Section B.2.2.4 of the CS[1] described an analysis of European Society for Blood and Marrow Transplantation (EBMT) Registry data, commissioned by the company to provide a comparison between partly published EBMT registry data on fludarabine/melphalan (Flu/Mel) and busulfan/cyclophosphamide (BU/CY) based conditioning treatment without TBI (Control EBMT) and fludarabine/treosulfan based conditioning of adult AML and MDS patients treated in MC-FludT.14/L Trial II (Treated chemotherapy [CT]) by using propensity score matching methods.[1] A summary of this analysis is provided in Table 4.13. In addition, a further registry study (CIBMTR),[16] was provided along with the company's response to clarification questions, but was not included in Document B of the CS[1] or utilised in the cost effectiveness modelling; a summary of this analysis, based on information taken from the study report,[16] has been added to Table 4.13.

The analysis comparing the EBMT registry data for control treatments (Flu/Mel and BU/CY without TBI) to fludarabine/melphalan from MC-FludT.14/L used propensity score matching to create similar patient groups for statistical comparisons. The EBMT data were retrospective whereas the trial data were prospective but single-arm, the analysis only included patients aged ≥50 years. There were some important statistically significant differences between the two groups, particularly for HCT-CI score, Karnofsky performance index, MUD donors, stem cell source, patient age and donor age which were all included in the matching model. Three different propensity score matching methods were used however the groups were not well-matched in terms of HCT-CI. Multivariable analyses were used to compare treatments after adjusting for HCT-CI, donor, stage, prognostic score and Karnofsky performance score with sensitivity analyses including other variables.

The CIBMTR analysis was also a comparison of retrospective registry data for comparator regimens and the fludarabine/melphalan arm from MC-FludT.14/L for the same outcomes plus EFS. However, this did not use matching methods and compared the two groups using a multivariable Cox proportional hazards regression with step-wise selection of variables. The OS model adjusted for performance score, HCT-CI, cytogenetics and donor-recipient sex match. Other variables, such as patient age which was significantly lower in the Cy/BU (myeloablative) group, were not adjusted for.

The ERG concludes that the EBMT comparison used appropriate statistical methods as it tried to match the treatment groups prior to comparing OS and the methods were suitable. The CIBMTR did not use an appropriate statistical method as there was no matching and only variables chosen by statistical selection were included in the final model, it was not based on clinical knowledge and differences between the treatment groups. However, the conclusions from both registry analyses are limited by the fact that these are comparisons of two groups from different studies (registry and prospective RCT) which have been matched using statistical methods and not randomisation. The analyses can only

include variables which have been measured in both studies, other important variables may differ between groups but cannot be included. The studies may also have important differences in patient inclusion criteria and data collection methods. As these conclusions are not based on comparisons from randomised controlled trials they should be interpreted with caution.

The ERG notes that the results of these analyses were not used in the cost effectiveness modelling.

Table 4.13: Summary of EBMT and CIBMTR registry analyses

Location	Europe	USA
Trial design	Medac commissioned registry-based retrospective study of partly published EBMT registry data base using propensity score matching methods	Medac commissioned comparison of CIBMTR registry data using multivariable Cox proportional hazards regression (OS and EFS) and Fine and Gray models (relapse, TRM, GvHD).
Number of patients	Control EBMT - [redacted] patients Treated CT - [redacted] patients	Control CIBMTR - [redacted] patients Treated CT - [redacted] patients
Eligibility criteria for participants	Patients from European Union countries aged between 50 and 70 with primary or secondary AML in CR or MDS patients (regardless of disease subtype (WHO) stage) Donor type HLA identical sibling/MRD or MUD with source of stem cells being bone marrow or peripheral blood Patients undergoing first alloHSCT between 2010 and 2016 with a Karnofsky score ≥60 and utilising Flu/Mel or BU/CY without TBI conditioning	Patients from the USA aged ≥50 years or aged 18 to 70 with HCT-CI score >2, with AML in complete remission (bone marrow blast <5 %) or MDS (bone marrow blast <20 %) Donor type HLA-matched sibling or HLA-matched adult unrelated donor (matched at the allele-level at HLA-A, -B, -C and -DRB1) Patients undergoing transplantation between 2009 and 2014 with a performance score ≥60 Conditioning regimens: RIC Bu/Flu; RIC Bu/Flu plus ATG; MAC Bu/Flu; MAC Bu/Flu plus ATG; MAC Bu/Cy GvHD prophylaxis: cyclosporine or tacrolimus with mycophenolate or methotrexate
Settings and locations where the data were collected	EBMT Registry database (patients from Registry from 2010 to 2016)	CIMBTR Registry database (patients from Registry from 2009 to 2014)
Trial drugs	Control EBMT - fludarabine/melphalan (Flu/Mel) and busulfan/cyclophosphamide (BU/CY) based conditioning treatment without TBI	Control CIMBTR - RIC Bu/Flu, RIC Bu/Flu plus ATG, MAC Bu/Flu, MAC Bu/Flu plus ATG and MAC Bu/Cy treatment without TBI

Location	Europe	USA
	Treated CT - fludarabine/treosulfan based conditioning of adult AML and MDS patients treated in MC-FludT.14/L Trial II	Treated CT - fludarabine/treosulfan based conditioning of adult AML and MDS patients treated in MC-FludT.14/L Trial II
Primary objectives	Key study objectives were to: Compare OS at two years after alloHSCT of fludarabine/melphalan and busulfan/cyclophosphamide based conditioning treatment with fludarabine/treosulfan based conditioning in MDS and AML patients separately. Compare cumulative incidence of relapse (RI) at two years after transplant of fludarabine/melphalan and busulfan/cyclophosphamide based conditioning treatment with fludarabine/treosulfan based conditioning in MDS and AML patients separately. Compare NRM at two years after transplant of fludarabine/melphalan and busulfan/cyclophosphamide based conditioning treatment with fludarabine/treosulfan based conditioning in MDS and AML patients separately.	To compare allogeneic transplantation outcomes for AML and MDS with full or less intensity busulfan-fludarabine regimens and busulfan-cyclophosphamide regimen in the United States (US) to a treosulfan-containing regimen in a medac-sponsored European (EU) clinical trial MC-FludT.14/L Trial II.
Summary of primary endpoint results	For MDS patients, treatment with fludarabine/treosulfan based conditioning lead to improved OS vs BuCy (HR [redacted] (95% CI [redacted]). Moreover, other comparisons (vs fludarabine + melphalan [FluMel]) were consistently in favour of fludarabine/treosulfan (HR < [redacted]). The differences were clinically relevant (over [redacted] difference for OS, up to [redacted]% difference for NRM, [redacted]% difference for RI at 2 years). For AML patients, OS [redacted] (95% CI [redacted]) vs FluMel; HR [redacted] (95% CI [redacted]) vs BuCy] is significantly improved, while NRM is significantly reduced (HR [redacted] (95% CI [redacted]) vs FluMel; HR [redacted] (95% CI 0.[redacted]) vs BuCy) with the use of fludarabine/treosulfan based conditioning compared to FluMel and BuCy.	For all patients (AML and MDS) and all comparators, treatment with RIC Treo/Flu was associated with improved OS: MAC Bu/Flu vs RIC Treo/Flu, HR [redacted] (95% CI: [redacted] to [redacted]) MAC Bu/Flu plus ATG vs RIC Treo/Flu, HR [redacted] (95% CI: [redacted] to [redacted]) MAC Bu/Cy vs RIC Treo/Flu, HR [redacted] (95% CI: [redacted] to [redacted]) RIC Bu/Flu vs RIC Treo/Flu, HR [redacted] (95% CI: [redacted] to [redacted]) RIC Bu/Flu plus ATG vs RIC Treo/Flu, HR [redacted] (95% CI: [redacted] to [redacted])

Location	Europe	USA

Based on Table 11 of the CS,[1] and text and Table 3 of the study report[16]
alloHSCT = allogeneic haematopoietic stem cell transplantation; AML = acute myeloid leukaemia; ATG = anti-thymocyte globulin ; Bu = busulfan; CI = confidence interval; CIBMTR = Center for International Blood and Marrow Transplant Research; CR = complete remission; CT = chemotherapy; Cy = cyclophosphamide; EBMT = European Society for Blood and Marrow Transplantation; EFS = event-free survival; Flu = fludarabine; GvHD = graft-versus-host disease; HLA = human leukocyte antigen; HR = hazard ratio; MAC = myeloablative conditioning; MDS = myelodysplastic syndrome; Mel = melphalan; MRD = Matched related donor; MUD = Matched unrelated donor; NRM = non-relapse mortality; OS = overall survival; RI = relapse incidence; RIC = reduced intensity conditioning; TBI = total body irradiation; TRM = transplantation-related mortality; WHO = World Health Organization

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

The CS did not include any indirect comparisons.

4.4 Critique of the indirect comparison and/or multiple treatment comparison

The CS did not include any indirect comparisons.

4.5 Additional work on clinical effectiveness undertaken by the ERG

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

4.6 Conclusions of the clinical effectiveness section

The evidence for treosulfan in combination with fludarabine for conditioning therapy as part of alloHSCT is based on a multi-centre, international RCT (MC-FludT.14/L Trial II) investigating patient-relevant outcomes, with follow-up to two years after transplantation.

However, the population included in MC-FludT.14/L Trial II is not representative of the full population defined in the scope, specifically, there is a lack of evidence about the effectiveness of treosulfan/fludarabine (Treo/Flu) conditioning regimens in people who are able to tolerate standard MAC regimens, in adults with malignancies other than AML or MDS, and in children.

Furthermore, the nature of the restricted population in MC-FludT.14/L Trial II (adults who were at increased risk and therefore not eligible for standard MAC conditioning regimens), and hence the choice of comparator (RIC Bu/Flu) means that the relative effectiveness and cost effectiveness of Treo/Flu has not been evaluated against the range of comparator regimens that would be relevant for the full population defined in the scope.

The CS[1] and the company's response to clarification questions[12] included two registry studies, EBMT[15] and CIBMTR.[16] These studies compared Treo/Flu-treated patients from MC-FludT.14/L Trial II with registry patients who had been treated with other conditioning regimens, and were commissioned by medac to provide information about the comparative effectiveness of Treo/Flu versus other conditioning regimens (see Sections 4.2.1 and 4.2.9 for details). Both studies included only registry patients who matched the reported criteria used in MC-FludT.14/L Trial II to define patients at increased risk and not eligible for standard high-intensity MAC conditioning regimens, however, the conditioning regimens received by registry patients in these studies included some high-intensity MAC regimens. The ERG therefore questions whether MC-FludT.14/L Trial II used any additional criteria to define increased risk patients, not eligible for standard high-intensity MAC conditioning regimens. The ERG is also unclear as to how similar the definition of patients at increased risk for standard conditioning therapies (i.e. not eligible for standard high-intensity MAC), used in MC-FludT.14/L Trial II, is to any such definition that would be generally applied in practice; it is important to establish a consistent definition in order to inform any recommendations for this population.

Irrespective of whether/the extent to which the registry studies can provide comparative effectiveness data for Treo/Flu versus relevant comparator regimens (including standard high-intensity MAC regimens where applicable), data from these studies have not been utilised in the cost effectiveness modelling. The ERG therefore considers that the evidence included in the submission is sufficient to support an assessment of the cost effectiveness of Treo/Flu versus RIC Bu/Flu in adults with AML or MDS, who are at increased risk for standard conditioning therapies (i.e. not eligible for standard high-

intensity MAC regimens), but is not sufficient to support an assessment of the cost effectiveness of Treo/Flu for the full scope population or versus any of the other comparators defined in the scope.

5. Cost effectiveness

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

The company conducted searches for economic, health related quality of life and cost/resource use evidence. A good range of databases, conference proceedings and additional resources were searched. The company submission and clarification response provided sufficient detail for the ERG to be able to appraise the searched conducted by the company.

5.1.1 Searches performed for cost effectiveness section

The following paragraphs contain summaries and critiques of all searches related to cost effectiveness presented in the company submission.

Appendix A of the CS details a systematic search of the literature used to identify cost effectiveness studies.[17] Separate searches in Embase were run for economic, cost/resource and health state utility studies. PubMed and Cochrane Library searches for economic, cost/resource and utility studies were combined with searches for RCTs. A range of conferences were also searched for cost effectiveness studies. A summary of the sources searched is provided in Table 5.1.

Table 5.1: Data sources for the cost effectiveness systematic review

Search strategy element	Resource	Host/Source	Date Range	Date searched
Electronic databases	Embase	Embase.com	2008-2019	11 February 2019. Update searches on 20 June 2019
	PubMed	www.ncbi.nlm.nih.gov/pubmed/	2008-2019	12 February 2019
	Cochrane Library (Cochrane Reviews, Cochrane Protocols, Clinical answers, Editorials and Special Collections)	http://onlinelibrary.wiley.com/cochranelibrary/search/	2008-2019	12 February 2019
Conference proceedings	ASH	2017 – host not reported 2018 - www.bloodjournal.org/page/ash-annual-meeting-abstracts?sso-checked=true	2017-2018	Not reported
	BSBMT	2017 - www.bsbmt.org/wp-content/uploads/2017/08/BSBMT-News-18d.pdf 2018 – not reported	2017-2018	Not reported

Search strategy element	Resource	Host/Source	Date Range	Date searched
	CIBMTR/ASBMT	2017 - www.bloodjournal.org/content/130/suppl_1?sso-checked=true 2018 – not reported	2017-2018	Not reported
	EHA	2017 – https://ehaweb.org/assets/Uploads/EHA22-Abstract-Book.pdf 2018 - https://journals.lww.com/hemasphere/Citation/2018/06001/23rd_Congress_of_the_European_Hematology.1.aspx	2017-2018	Not reported
	EBMT	https://www.ebmt.org/annual-meeting>abstracts	2017-2018	Not reported
ASBMT = American Society for Blood and Marrow Transplantation; ASH = American Society of Haematology; BSBMT = British Society of Blood and Marrow Transplantation; CIBMTR = Centre for International Blood and Marrow Transplant Research; EBMT = European Society for Blood and Marrow Transplantation; EHA = European Hematology Association;				

ERG comment:

- The ERG considers the database searches and methodology reported in the CS to support the systematic review of cost effectiveness to be comprehensive, transparent, reproducible and fit for purpose.
- Validated search filters were not applied and would have improved the sensitivity and comprehensiveness of searches for economic, cost and health state utility studies.
- An additional search of conference proceedings was also undertaken to find unpublished and ongoing studies on cost effectiveness and utilities.

5.1.2 Inclusion/exclusion criteria

Detailed inclusion/exclusion criteria which were applied to the studies identified in the cost effectiveness searches were provided by the company in Table 1 of Appendix D.[17] Inclusion/exclusion criteria were based on the PICOS criteria, to identify the population and disease, interventions, comparators, outcomes, and study designs of interest, as well as publication types, publication dates and language. All studies published prior to 2008 (prior to 2017 for conferences) and all non-English language papers were excluded.

During the full text review, the eligibility criteria used in the first pass screening were slightly modified to include publications from phase II studies only from treosulfan (Phase II studies from the comparator drugs were excluded) Studies reporting HRQoL data (flagged in the first pass screening) based on

generic preference based quality of life (QoL) tools or those that could be mapped to EQ-5D were also included. For studies reporting cost and resource use data, only studies from five European countries were considered (France, Germany, Italy, Spain and the UK). For cost effectiveness analyses, only studies focusing on the relevant intervention or listed comparators were considered for inclusion.

ERG comment: The ERG was concerned that the language limitation of only English language publications may have introduced potential language bias. Current best practice states that "*Whenever possible review authors should attempt to identify and assess for eligibility all possibly relevant reports of trials irrespective of language of publication*".[39] The geographical criterion for cost/resource use data (only including data from France, Germany, Italy, Spain and the UK) was considered restrictive by the ERG, as information from other countries could also provide relevant information and adjustments to UK costs could also be made to data from countries outside of these five, in the same way as it would be for the non-UK studies from included countries.

5.1.3 Identified studies

The reporting of the number of records returned by the SLR searches was not separated into clinical and cost effectiveness searches. In total, the electronic databases searched provided 13,091 unique references for screening. Of these, 3,881 were excluded based on type and year of publication and 8,347 references were excluded based on other exclusion criteria (details provided in Figure 1 of Appendix D).[17] Eight hundred and sixty-three studies were assessed at full text. During the full text review, 803 publications were excluded. Study design (n=346), intervention not of interest (n=193), population (n=102) and publication type (n=66), were the major reason for exclusion of the citations. However, it is unclear how many of these were related to each of the different evidence parameter searches. The full text review resulted in selection of 60 studies for inclusion.

A total of 7,429 hits were identified in the manual congress searches (details provided in Table 3 of Appendix D).[17] Title screening resulted in shortlisting of 294 relevant abstracts. Of these, one was a duplicate of a reference identified in the electronic searches and 286 abstracts were excluded for having limited data or inappropriate study design. Seven conference abstracts were selected for inclusion in the review.

The final list of 64 references included 28 clinical studies, and 36 cost effectiveness, HRQoL or cost/resource studies. Three studies describing cost effectiveness analysis of conditioning therapies were identified, while only one study investigating resource use in this indication was included. The systematic review of health state utility values identified, included a total of 32 publications, 15 of which provided utility values for different health states in patients undergoing HSCT, and 15 which provided data on QoL measures that could be mapped to EQ-5D utilities. Seventeen additional publications were selected that reported QoL data using generic preference-based tools that could be mapped to utility values and two studies included, developed mapping algorithms. Details of all included studies from the cost effectiveness searches are provided in Sections B.1.5.3 – B.1.5.5 of Appendix D.[17]

ERG comment: It would have been useful to see separate PRISMA diagrams for the cost effectiveness searches in order to have a better idea of how certain restrictive exclusion criteria, such as the geographical restriction on cost and resource use data impacted the results of the review.

5.1.4 Interpretation of the review

The review was generally well reported and identified a range of cost effectiveness evidence relevant to the indication and population and useful for the cost effectiveness analysis.

5.2 Summary and critique of company's submitted economic evaluation by the ERG

A summary of the economic evaluation conducted by the company is presented in Table 5.2.

Table 5.2: Summary of the company submission economic evaluation

	Approach	Source/Justification	Signpost (location in ERG report)
Model	A partitioned-survival model was considered to evaluate the cost effectiveness of treosulfan with fludarabine as a conditioning treatment for malignant disease (AML or MDS) prior to allogeneic haematopoietic stem cell transplantation in adults.	The modelling approach was deemed appropriate considering the data available. It was in line with TA523[40] and TA552[41], two recent technology appraisals on AML.	Section 5.2.2
States and events	After receiving an HSCT, all patients start in the remission health state of the model. From there, patients can either relapse or die. Transitions between health states are determined by EFS and OS curves obtained by fitting statistical models to the survival data from the MC-FludT.14/L Trial II.		Section 5.2.2
Comparators	The comparator is busulfan with fludarabine.	This is the comparator in MC-FludT.14/L Trial II.	Section 5.2.4
Natural history	Autologous HSCT uses the patient's own cells, allogeneic HSCT uses cells from a donor.[2] Both HSCTs are potentially curative for a host of different diseases. *"Allogeneic HSCT is recommended for acute and chronic leukaemias, myelo-dysplastic syndromes (MDS), HL, NHL, and MM"*.[2] Allogeneic HSCT involves a certainty of rejection unless the patient is first treated with myeloablative conditioning. Standard myeloablative conditioning (MAC) regimens generally lead to low relapse rates, but have high treatment-related toxicity and transplant-related mortality (TRM). Non-myeloablative (NMA) conditioning and reduced intensity conditioning (RIC) was developed for patients such as the elderly and those with comorbidities where myeloablative conditioning is not		Section 2.1

	Approach	Source/Justification	Signpost (location in ERG report)
	considered optimal and to minimise treatment-related toxicity, non-relapse mortality (NRM) and transplant-related mortality (TRM).[1] However, lower dose intensity conditioning has higher rate of relapse after allogeneic HSCT.[3]		
Treatment effectiveness	Treatment effectiveness was based on the results from MC-FludT.14/L Trial II.		Section 5.2.6
Adverse events	The following adverse events (AEs) were taken into account: Extensive chronic graft versus host disease (cGvHD), Stage III/IV acute graft versus host disease (aGvHD), oral mucositis, nausea, diarrhoea, vomiting, gamma glutamyl transferase increased, alanine aminotransferase increased, aspartate aminotransferase increased, blood bilirubin increased, Investigations (other), febrile neutropenia, sepsis, lung infection, anorexia, syncope, and maculopapular rash.	The effects of AEs are captured by applying a utility decrement over a stated time period based on data from the phase III clinical trial, previous NICE technology appraisals, and validation by clinical expert opinion.	Section 5.2.7
Health related QoL	The company searched the literature and previous NICE technology appraisals for health state utility values and AE disutilities. Values were chosen to best reflect the NICE reference case and population. Utility values were adjusted for age and sex.	A large number of potential sources for health state utility values were considered. No UK EQ-5D values in the correct population were identified and therefore in the base-case, mapping was used to map from QLQ-C30 data to EQ-5D values using established mapping algorithms. Utility decrement values were selected from prior NICE appraisals or published literature.	Section 5.2.8
Resource utilisation and costs	The model included treatment costs (incl. concomitant medication), HSCT procedure costs, costs associated with HSCT after the procedure itself (incl. monitoring), costs for relapsed and progressed	Unit costs were obtained from the PSSRU 2018,[42] and NHS reference costs.[20] Drug costs were taken from the BNF. Transfusion costs were sourced from	Section 5.2.9

	Approach	Source/Justification	Signpost (location in ERG report)
	disease (incl. inpatient days, contacts with various specialists and nurses, monitoring, blood transfusions, and treatment costs), mortality costs, and costs due to adverse events.	TA399[43]. Treatment costs for patients with relapsed AML and MDS are based on key opinion leader input, with dosages and treatment sequences for each regimen sourced from protocols used by the South East London Cancer Network. Incidence of AEs are based on data from MC-FLUDT14/L Trial III. This applies also to the durations of AEs, except for the durations of cGvHD and aGvHD that were based on TA545,[44] and for oral mucositis, nausea, diarrhoea, vomiting, sepsis, and lung infection the durations are assumed to be the average for the associated HRG codes.	
Discount rates	A 3.5% discount rate was used for both costs and effects.	According to NICE reference case.	Section 5.2.5
Sensitivity analysis	Probabilistic and one-way sensitivity analysis.	According to NICE reference case.	Section 6.2.1

AE = adverse event; aGvHD = acute graft-versus-host disease; AML = acute myeloid leukaemia; BNF = British National Formulary; cGvHD = chronic graft-versus-host disease; EFS = event-free survival; EQ-5D = European Quality of Life-5 Dimensions; ERG = Evidence Review Group; HRG = healthcare resource group; HRQoL = health related quality of life; HSCT = haematopoietic stem cell transplantation; MDS = myelodysplastic syndrome; NHS = National Health Service; NICE = National Institute for Health and Care Excellence; OS = overall survival; PSS = Personal Social Services; PSSRU = Personal Social Services Research Unit; QALY = quality adjusted life year; QLQ-C30 = Quality of Life Core Questionnaire C30; TA = technology appraisal

5.2.1 NICE reference case checklist (TABLE ONLY)

Table 5.3: NICE reference case checklist

Element of health technology assessment	Reference case	ERG comment on company submission
Perspective on outcomes	All direct health effects, whether for patients or, when relevant, carers	Outcome measures included in the CS: • EFS (event-free survival) • OS (overall survival) • Adverse effects of treatment • Health-related quality of life
Perspective on costs	NHS and PSS	According to NICE reference case
Type of economic evaluation	Cost utility analysis with fully incremental analysis	According to NICE reference case
Time horizon	Long enough to reflect all important differences in costs or outcomes between the technologies being compared	Time horizon can be considered lifetime (model time horizon of 40 years for cohort with mean age of 60 years at baseline)
Synthesis of evidence on health effects	Based on systematic review	Systematic literature reviews were conducted for relevant cost effectiveness studies, and studies on HRQoL, cost and resource utilisation for the population in this assessment.
Measuring and valuing health effects	Health effects should be expressed in QALYs. The EQ-5D is the preferred measure of health-related quality of life in adults.	According to NICE reference case. No UK EQ-5D was identified in this population, so mapping was used to convert HRQoL data from other measures to UK EQ-5D utility values
Source of data for measurement of health-related quality of life	Reported directly by patients and/or carers	Utility data were not available from the trial and were searched for in the literature. Many potential sources of patient reported HRQoL data were identified, however no UK EQ-5D was found in this population.
Source of preference data for valuation of changes in health-related quality of life	Representative sample of the UK population	In the absence of UK EQ-5D data the company used established and appropriate mapping algorithms to map to UK EQ-5D utility values, as preferred by NICE in the situation where no UK EQ-5D values are available.

Element of health technology assessment	Reference case	ERG comment on company submission
Equity considerations	An additional QALY has the same weight regardless of the other characteristics of the individuals receiving the health benefit	According to NICE reference case
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	According to NICE reference case
Discounting	The same annual rate for both costs and health effects (currently 3.5%)	According to NICE reference case

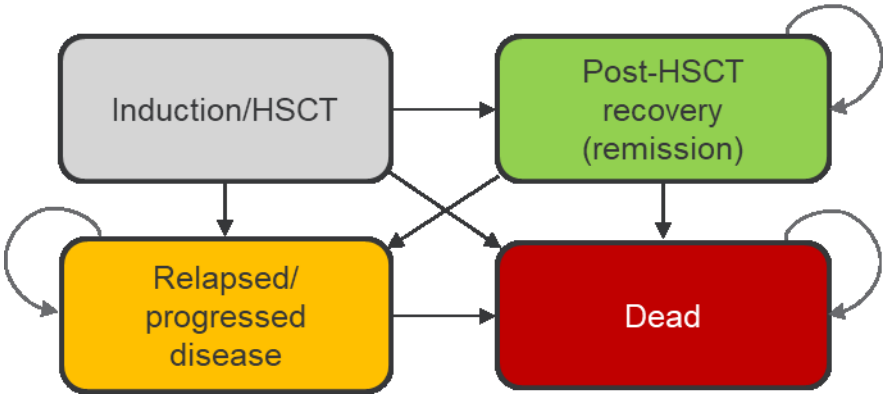
CS = company submission; EQ-5D = European Quality of Life-5 Dimensions; HRQoL = health-related quality of life; NHS = National Health Service; NICE = National Institute for Health and Care Excellence; OS = overall survival; PSS = personal social services; QALY = quality-adjusted life year; TA = technology appraisal

5.2.2 Model structure

A *de novo* partitioned-survival model to evaluate the cost effectiveness of treosulfan with fludarabine compared to busulfan with fludarabine as a conditioning treatment for malignant disease (AML or MDS) prior to allogeneic haematopoietic stem cell transplantation in adults, was developed by the company. The modelling approach was deemed appropriate by the company considering the clinical data available, and it was in line with TA523[40] and TA552[41], two recent technology appraisals on AML.

The simulation assumes that all patients have received an HSCT. Therefore, all patients start in the induction/HSCT health state. From there, patients can enter post-HSCT recovery (remission), relapse/progress or die. Transitions between health states are determined by event-free survival (EFS) and overall survival (OS) curves obtained by fitting statistical models to the survival data from the MC-FludT.14/L Trial II. The structure of the model is shown in Figure 5.1. Health states costs and utilities are used to calculate total costs and total quality-adjusted life years (QALYs) over a lifetime time horizon. A 28-day model cycle was assumed and half-cycle correction was applied to the cost and QALY calculations.

Figure 5.1: Model structure



Based on Figure 11 of the CS.[1]
CS = company submission; HSCT = haematopoietic stem cell transplantation

ERG comment: The modelling approach considered by the company is in line with those in two recent NICE technology appraisals on AML (TA523[40] and TA552[41]). The ERG considers this approach appropriate for the decision problem at hand.

On page 52 of the CS, EFS was defined as follows: "*Event-free survival within 2 years after transplantation was defined as the primary endpoint of the trial. EFS was measured from time of end of HSCT (= Day 0) to time of event. Events were defined as relapse of disease, graft failure or death (whatever occurred first)*".[1] In the request for clarification (question C3) the ERG asked the company to explain why graft failure was not modelled as a health state.[12] In their response, the company indicated that since graft failures occurred in very few patients in MC-FludT.14/L Trial II, extrapolating long-term survival curves for graft failure patients would be unreliable. Furthermore, the company did not identify any data indicating differences between relapse/progression patients and graft failure patients in terms of either costs or quality of life. Therefore, the company included graft failure as an adverse "event" in terms of EFS. Thus, in the model, graft failure patients were considered in terms of survival but were not differentiated from relapse/progression patients in terms of costs and non-survival outcomes. The ERG agrees with this approach.

5.2.3 Population

The population considered in the economic evaluation is the same as in the treosulfan trial MC-FludT.14/L Trial II: *"patients aged ≥ 50 years at transplant and/or with a Haematopoietic Cell Transplantation-Comorbidity Index (HCT-CI) score >2 (indicating an intermediate to high risk of co-morbidities which influence NRM)"*.[1]

ERG comment: As previously discussed in Section 4.6 of this report, the population considered by the company in the economic evaluation is not in line with the population defined in the final scope from NICE: *"adults, children and young people with malignant disease that is in remission before allogenic haematopoietic stem cell transplantation"*.[8] It is thus uncertain whether the results of the economic analyses can be generalised to people who are able to tolerate standard MAC regimens, in adults with malignancies other than AML or MDS, and in children.

The study population is reflective of the UK population covered by the indication, see Section 4.2.3 for further details.

5.2.4 Interventions and comparators

The intervention considered in the cost effectiveness analyses is treosulfan in combination with fludarabine as part of conditioning treatment prior to allogeneic haematopoietic stem cell transplantation (alloHSCT) in adult patients with malignant diseases (AML or MDS). Administration of treosulfan should be supervised by a physician with experience in conditioning treatment. The recommended dose and administration schedule are the same as in the treosulfan trial MC-FludT.14/L Trial II (see Section 3.2 of this report for further details).

The comparator considered in the cost effectiveness analyses is also the same as in the treosulfan trial MC-FludT.14/L Trial II: busulfan in combination with fludarabine as part of conditioning treatment prior to alloHSCT. For details on the recommended dose and administration schedule see Table 4.3.

ERG comment: The comparator considered by the company in the economic evaluation is just one of the comparator treatments defined in the final scope from NICE.[8] The full list of comparators included in the scope can be seen in Table 3.1. As previously discussed in Section 4.6 of this report, no other comparators listed in the scope were included in any economic analysis. In line with Section 5.2.3, it is uncertain whether treosulfan would be cost effective in people who are able to tolerate standard MAC regimens, i.e. in adults with malignancies other than AML or MDS, and in children.

5.2.5 Perspective, time horizon and discounting

The economic analyses were conducted from the perspective of the NHS and Personal Social Services (PSS). A 40-year horizon was assumed in the model, which is long enough to be considered as lifetime. Costs and QALYs were discounted at 3.5% per annum according to the NICE method guidance.[45]

5.2.6 Treatment effectiveness and extrapolation

5.2.6.1 Survival analysis

General approach

During MC-FludT.14/L Trial II, the company collected survival data up to 1,586 days (the duration of the clinical trial). In order to extrapolate these data beyond the end of the clinical trial, several survival regression models were fitted to the clinical data. The following approaches were considered by the company for both OS and EFS:

- Proportional hazards models
- Standard parametric models (e.g. exponential, Weibull, Gompertz, etc.)
- Mixture-cure models (MCM) / non-mixture-cure models (NMCM)
- Flexible spline models

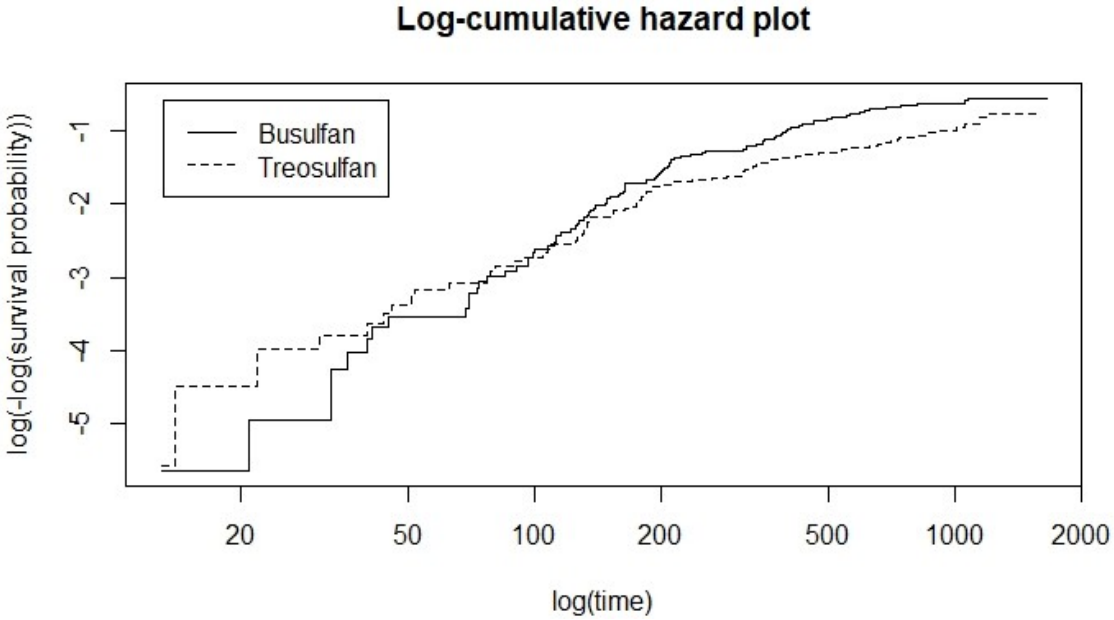
According to the company, proportional hazards could not be assumed based on log-log plots and flexible spline models were not considered due to concerns with over-fitting the observed data. Therefore, standard parametric models and MCMs/NMCMs were further explored by the company.

ERG comment: Statements made by the company in the CS over proportional hazards and flexible spline models were not supported by any evidence.

Log-log plots for both OS and EFS were provided by the company in response to the clarification question C9.a and presented below in Figure 5.2 and 5.3 for OS and EFS, respectively.[12] Since for both

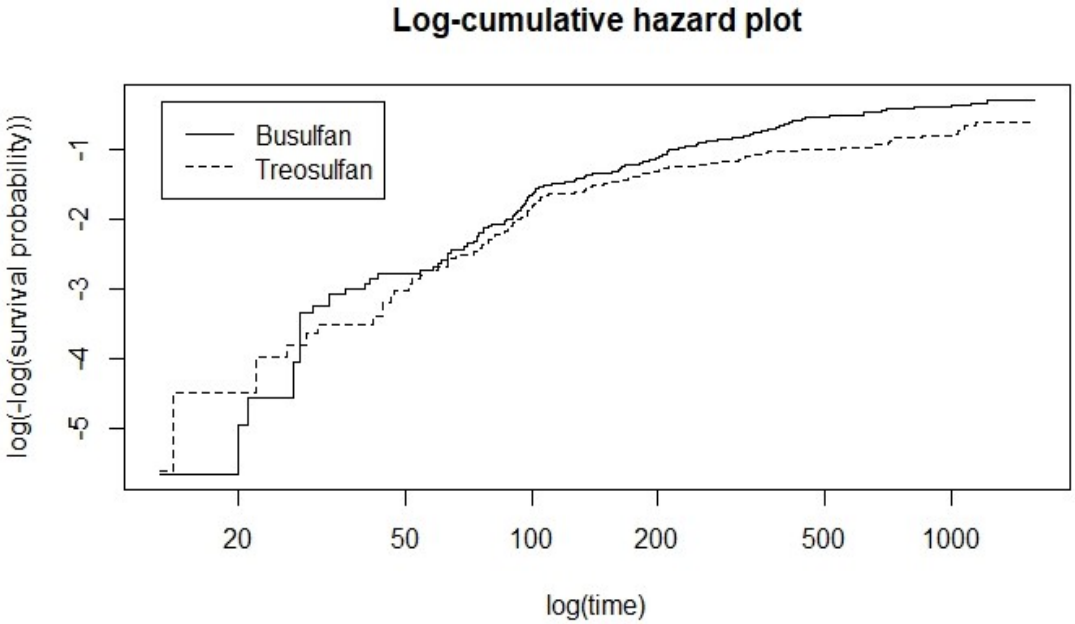
OS and EFS, the curves for treosulfan and busulfan are clearly non-parallel and they even overlap, the ERG agrees with the company in that assuming proportional hazards does not seem appropriate.

Figure 5.2: OS log-log cumulative hazards plot



Based on Figure 4 of the response to request for clarification[12]
OS = overall survival

Figure 5.3: EFS log-log cumulative hazards plot



Based on Figure 5 of the response to the request for clarification [12]
EFS = event-free survival

Regarding flexible spline models, the company indicated in response to clarification question C9.h that "*no formal statistical analyses were performed to determine whether the flexible spline models over-fit. The parametric survival models and mixture-cure models appeared to fit the data sufficiently well, so the flexible spline approach was not pursued further given that it also did not appear to produce*

significant improvements in fit according to AIC statistics and visual inspection".[12] This answer seems to suggest that the flexible spline models were initially considered but not included in the final analyses due to AIC and visual inspection assessments. Furthermore, the company mentioned that another disadvantage of flexible spline models is that they are more complex and were not used in the base-case analysis of any previous AML single technology appraisals (STAs). While this might be the case, it remains unclear why the company indicated in the CS that these models were not considered due to concerns with over-fitting the observed data. The ERG would like to emphasise that it is not suggesting using flexible spline models, but simply pointing out an inconsistency. Should these models be considered in the economic model, it is unclear how these would impact the results.

Fitting and selection of survival models

The full survival dataset was used for fitting (i.e. no data cuts were considered). Analyses were carried out for the pooled AML and MDS cohorts but also for the AML and MDS subgroups separately. Analyses were stratified by treatment arm (because proportional hazards were not assumed). The same approach was taken for OS and EFS. Goodness of fit was assessed for all standard parametric models (exponential, Weibull, lognormal, log-logistic, Gompertz and gamma) and for Weibull and lognormal MCMs and NMCMs.

The CS states that model selection followed the recommendations of the NICE Decision Support Unit (DSU) guidelines and was based on the following criteria:[46]

- Akaike information criterion (AIC) and Bayesian information criterion (BIC)
- Visual inspection
- Key Opinion Leader (KOL) feedback

However, the CS states subsequently that models were primarily selected based on (the lowest) AIC and BIC. Furthermore, the company assumed the same type of model for treosulfan and busulfan according to the advice in NICE DSU 14.[46] The AIC and BIC values for each parametric model considered by the company, for both treosulfan and busulfan, OS and EFS, are shown in Tables 5.4 to 5.7. Based on these values, the company chose a log-normal NMCM for all analyses, since this model had the lowest AIC for EFS in the pooled AML + MDS population for both arms, the lowest AIC for OS in the pooled AML + MDS population the busulfan arm, and the second best fit for OS for the AML + MDS patients in the treosulfan arm. Of the standard survival models, the company considered that the gamma produced the most consistent estimates and, because of this, it was explored in scenario analysis.

Table 5.4: AIC and BIC for the models fit to the OS data in the treosulfan arm

Population	AML + MDS		AML		MDS	
Regression model	**AIC**	**BIC**	**AIC**	**BIC**	**AIC**	**BIC**
Exponential	956.20	959.42	1421.44	1425.04	466.25	468.70
Weibull	952.25	958.69	1412.05	1419.25	462.86	467.75
Lognormal	946.78	953.22	1403.80	**1411.00**	459.94	**464.83**
Log-logistic	950.17	956.61	1409.14	1416.34	461.98	466.87
Gompertz	**946.17**	**952.61**	1404.89	1412.08	462.05	466.94
Gamma	946.90	956.56	**1402.15**	1412.95	**459.77**	467.10
MCM Weibull	948.18	957.83	1407.53	1418.30	463.79	471.08
MCM log-normal	946.39	956.04	1402.78	1413.56	461.11	468.40

Population	AML + MDS		AML		MDS	
Regression model	**AIC**	**BIC**	**AIC**	**BIC**	**AIC**	**BIC**
NMCM Weibull	947.82	957.47	1406.88	1417.66	463.49	470.79
NMCM Log-normal	946.34	955.99	1402.74	1413.51	461.09	468.38

Based on Table 37 of the CS[1]

Bold figures indicate that the model had the lowest AIC/BIC estimate for the associated patient population.

AIC = Akaike information criterion; AML = acute myeloid leukaemia; BIC = Bayesian information criterion; CS = company submission; MCM = mixture-cure model; MDS = myelodysplastic syndrome; NMCM = non-mixture-cure model; OS = overall survival

Table 5.5: AIC and BIC for the models fit to the OS data in the busulfan arm

Population	AML + MDS		AML		MDS	
Regression model	**AIC**	**BIC**	**AIC**	**BIC**	**AIC**	**BIC**
Exponential	1039.94	1043.06	1880.66	1884.31	840.45	843.20
Weibull	1039.12	1045.37	1872.45	1879.74	834.41	839.90
Lognormal	1026.82	1033.07	1852.10	1859.39	825.37	830.86
Log-logistic	1033.37	1039.61	1861.97	1869.26	829.36	834.85
Gompertz	1025.49	1031.73	1844.89	1852.18	820.58	**826.07**
Gamma	1016.39	1025.76	1840.41	1851.34	821.70	829.94
MCM Weibull	1017.60	1026.97	1834.22	1845.16	818.80	827.03
MCM Log-normal	1010.96	1020.33	**1831.34**	**1842.27**	818.87	827.11
NMCM Weibull	1016.60	1025.97	1833.20	1844.13	**818.61**	826.85
NMCM Log-normal	**1010.76**	**1020.13**	1831.56	1842.49	819.10	827.34

Based on Table 39 of the CS[1]

Bold figures indicate that the model had the lowest AIC/BIC estimate for the associated patient population.

AIC = Akaike information criterion; AML = acute myeloid leukaemia; BIC = Bayesian information criterion; CS = company submission; MCM = mixture-cure model; MDS = myelodysplastic syndrome; NMCM = non-mixture-cure model; OS = overall survival

Table 5.6: AIC and BIC for the models fit to the EFS data in the treosulfan arm

Population	AML + MDS		AML		MDS	
Regression model	**AIC**	**BIC**	**AIC**	**BIC**	**AIC**	**BIC**
Exponential	1141.10	1144.32	1647.18	1650.78	507.98	510.42
Weibull	1119.22	1125.66	1616.80	1623.99	501.50	506.39
Lognormal	1106.67	1113.12	1600.05	1607.34	497.36	502.25
Log-logistic	1114.05	1120.49	1610.12	1617.41	500.05	504.93
Gompertz	1099.75	1106.19	1595.97	1603.26	499.50	504.38
Gamma	1091.17	1100.83	**1580.21**	**1591.15**	**494.80**	**502.13**
MCM Weibull	1103.08	1112.73	1599.83	1610.60	501.29	508.59
MCM Log-normal	1091.38	1101.02	1583.88	1594.65	497.18	504.47
NMCM Weibull	1101.69	1111.34	1597.87	1608.64	500.77	508.06
NMCM Log-normal	**1090.70**	**1100.34**	1582.98	1593.75	497.00	504.29

Based on Table 38 of the CS[1]

Bold figures indicate that the model had the lowest AIC/BIC estimate for the associated patient population.

Population	AML + MDS		AML		MDS	
Regression model	**AIC**	**BIC**	**AIC**	**BIC**	**AIC**	**BIC**

AIC = Akaike information criterion; AML = acute myeloid leukaemia; BIC = Bayesian information criterion; CS = company submission; EFS = event-free survival; MCM = mixture-cure model; MDS = myelodysplastic syndrome; NMCM = non-mixture-cure model

Table 5.7: AIC and BIC for the models fit to the EFS data in the busulfan arm

Population	AML + MDS		AML		MDS	
Regression model	**AIC**	**BIC**	**AIC**	**BIC**	**AIC**	**BIC**
Exponential	1293.49	1296.62	2206.65	2210.29	915.02	917.77
Weibull	1281.64	1287.89	2157.49	2185.62	900.24	905.73
Lognormal	1265.20	1271.45	2150.29	2157.49	888.28	893.77
Log-logistic	1272.44	1278.69	2162.33	2169.62	893.23	898.72
Gompertz	1260.34	1266.59	2137.10	2144.30	879.65	885.14
Gamma	1250.91	1260.28	2125.60	2136.40	878.46	886.70
MCM Weibull	1261.35	1270.72	2136.63	2147.57	879.85	888.09
MCM Lognormal	1250.60	1259.97	2122.69	2133.63	**876.56**	**884.80**
NMCM Weibull	1259.46	1268.83	2134.13	2145.07	879.16	887.39
NMCM Lognormal	**1250.23**	**1259.60**	**2122.51**	**2133.45**	876.69	884.92

Based on Table 40 of the CS[1]

Bold figures indicate that the model had the lowest AIC/BIC estimate for the associated patient population.

AIC = Akaike information criterion; AML = acute myeloid leukaemia; BIC = Bayesian information criterion; CS = company submission; EFS = event-free survival; MCM = mixture-cure model; MDS = myelodysplastic syndrome; NMCM = non-mixture-cure model

ERG comment: Despite mentioning that model selection followed the recommendations of the DSU guidelines,[46] and indicating AIC, BIC, visual inspection and KOL feedback as specific criteria for assessing goodness of fit, it seems that the final selection of parametric models was made solely based on AIC/BIC. Assessments based on visual inspection and KOL feedback were not reported in the CS. In response to the clarification question C9.i the company confirmed that "*AIC was ultimately the deciding factor in selecting the parametric curves used*".[12] The company concluded that while "*MCM Weibull and NMCM Weibull models also produced reasonable fits based on visual inspection, NMCM lognormal and MCM lognormal consistently produced the lowest AIC estimates across EFS and OS curves for treosulfan and busulfan for the overall population. Limited differentiation between NMCM lognormal and MCM lognormal was possible on the basis of AIC and visual inspection, and as such NMCM lognormal estimates were selected on the basis of AIC*" (see response to the clarification question C9.i).[12] The gamma distribution was chosen for scenario analysis.

Feedback from two UK based clinical experts was provided by the company in response to the clarification question C2.[12] However, this feedback was mostly regarding the assumptions around long-term mortality and the use of a five-year cure point, which will be discussed in the next section of this report. One of the experts raised an initial concern around "*the slight overestimation of patients in the relapse health state in the long-term*" but considered that the curves used in the base-case analysis (NMCM lognormal) were still plausible.[12] Therefore, clinical opinion about the plausibility of the other parametric survival models was not presented by the company.

In response to the clarification question C9.i, the company assessed goodness of fit based on visual inspection of the survival curves as follows:[12]

- NMCM/MCM lognormal produced the best fit
- NMCM/MCM Weibull produced a reasonable fit

While this might be the case, the ERG considers that a more detailed assessment should have been provided. After visual inspection of the curves in the electronic model, the ERG concluded that:

- Exponential, Weibull, lognormal and log-logistic curves do not seem to fit EFS or OS data for treosulfan and busulfan, and should not be considered as potential candidates for curve extrapolation.
- For the other parametric curves considered by the company, the visual fit assessment was not so clear and varied with the type of survival data and treatment arm:
    - Treosulfan EFS: NMCM/MCM lognormal, NMCM/MCM Weibull and Gompertz curves seem to provide a similar fit and to overestimate EFS at the tail of the Kaplan-Meier (KM) curve. The gamma distribution results in a less optimistic extrapolation for treosulfan and might be seen as a more conservative approach.
    - Busulfan EFS: Gompertz, NMCM/MCM lognormal seem to reasonably fit well the tail of the KM curve. NMCM/MCM Weibull provide a similar overestimation of EFS at the tail of the KM curve. The gamma provides a poor fit (large underestimation of EFS for busulfan).
    - Treosulfan OS: all the models provide a similar reasonably good fit.
    - Busulfan OS: NMCM/MCM Weibull seem to reasonably fit well the tail of the KM curve. NMCM/MCM lognormal provide a similar underestimation of OS at the tail of the KM curve. The Gompertz, and specially the gamma distribution provide a poor visual fit (underestimation of OS for busulfan).

The ERG considers that, while the selection of the NMCM lognormal distribution for the base-case might seem reasonable, it should be emphasised that this distribution seems to overestimate treosulfan EFS and to underestimate busulfan OS at the tail of the KM curves. Moreover, the selection of the gamma distribution as an alternative extrapolation is not appropriate since it only seems to fit well treosulfan data. For busulfan, the gamma distribution underestimates quite considerably both EFS and OS at the tail of the KM curves. For that reason, the ERG conducted an overall goodness-of-fit assessment based on all the information presented by the company (visual fit to KM, KOL feedback, AIC and BIC goodness-of-fit statistics) either in the main submission document, in the appendices or in the response to the request for clarification. A summary for the AML + MDS pooled population is presented in Table 5.8 to 5.11. The following assumptions were made by the ERG while assessing the goodness-of-fit of the parametric distributions:

- All methods used to assess goodness-of-fit (visual fit to KM curves, KOL feedback, AIC and BIC goodness-of-fit statistics) were considered equally important.
- However, given the limited KOL feedback provided (i.e. the plausibility of the NMCM lognormal curves only), this criterion was not used.
- Based on Burnham and Anderson rule of thumb,[47] it was considered that a difference in AIC less than 4 with respect to the AIC for the model with the lowest AIC was appropriate, between 4 and 7 was neutral, and larger than 10 was inappropriate.

- Based on Raftery rule of thumb,[48] it was considered that a difference in BIC larger than 10 with respect to the BIC for the model with the lowest BIC was inappropriate.

Based on the assessment summarised in Tables 5.8 to 5.11, the ERG considered that the best candidate distributions were the following:

- Treosulfan EFS: Gamma, NMCM/MCM lognormal
- Busulfan EFS: NMCM/MCM lognormal, Gompertz
- Treosulfan OS: Gompertz, Gamma, NMCM/MCM Weibull, NMCM/MCM lognormal, lognormal
- Busulfan OS: NMCM/MCM Weibull, NMCM/MCM Lognormal

By further assuming the same type of model for treosulfan and busulfan, as recommended in NICE DSU 14,[46] the ERG also chose the NMCM lognormal distribution to model EFS for its preferred base-case. The MCM lognormal distribution for EFS was explored in a scenario analysis. However, as previously mentioned, these two distributions seem to overestimate treosulfan EFS. The only way to avoid this issue, is by choosing different parametric distributions to model EFS for treosulfan and busulfan, which is contrary to the advice in DSU 14.[46] In DSU 14 it is mentioned that if "*different types of model seem appropriate for each treatment arm this should be justified using clinical expert judgement, biological plausibility, and robust statistical analysis*".[46] The ERG cannot justify the choice of different parametric distributions per treatment based on the criteria mentioned in DSU 14.[46] However, the ERG considers that choosing different types of curve for each treatment arm can be useful to estimate the size of the bias in the base-case. Therefore, the ERG conducted an additional scenario assuming a gamma distribution for treosulfan EFS and an NMCM lognormal distribution (as in the base-case) for busulfan EFS.

For OS, all four NMCM/MCM Weibull, NMCM/MCM lognormal seem to be appropriate choices. Based on visual fit, the both NMCM and MCM lognormal seem to underestimate OS for busulfan despite having a lower AIC than the NMCM/MCM Weibull. The NMCM Weibull was chose by the ERG for its preferred base-case since it had a lower AIC value than the MCM Weibull. The other three distributions were explored in scenario analyses.

Assessing goodness of fit based on AIC or BIC solely can be misleading. The gamma distribution in this case is a good example of this. The gamma distribution performed quite well in terms of AIC and BIC and, for that reason, it was explored by the company in scenario analysis. However, visual inspection of the curves indicated that the gamma distribution is considerably biased against busulfan (both EFS and OS) and, therefore, the results of that scenario analysis can be misleading.

Note finally that the ERG goodness of fit assessment presented in this section was done for the pooled AML + MDS population. A similar assessment was also done for the AML and MDS subgroups separately. These are presented in Appendix 1.

Table 5.8: Overall goodness-of-fit assessment by the ERG: treosulfan EFS in the AML + MDS population

Parametric model	Visual fit with KM	AIC	BIC	Overall assessment [score]
Exponential	☹	☹	☹	☹ [1]
Weibull	☹	☹	☹	☹ [1]
Lognormal	☹	☹	☹	☹ [1]
Log-logistic	☹	☹	☹	☹ [1]
Gompertz	😐	☹	☺	😐 [2]
Gamma	☺	☺	☺	☺ [3]
MCM Weibull	😐	☹	☹	☹ [1.3]
MCM Lognormal	😐	☺	☺	☺ [2.7]
NMCM Weibull	😐	☹	☹	☹ [1.3]
NMCM Lognormal	😐	☺	☺	☺ [2.7]

Visual fit with KM: Based on ERG assessment.
AIC: Define $\Delta i = AICi - AICmin$, where AICi is the AIC of the i-th model, and AICmin is the lowest AIC, obtained for NMCM lognormal: ☺ = $\Delta i < 4$; 😐 = $4 < \Delta i < 7$; ☹ = $\Delta i > 10$
BIC: Define $\Delta i = BICi - BICmin$, where BICi is the BIC of the i-th model, and BICmin is the lowest BIC, obtained for NMCM lognormal: ☺ = $\Delta i < 10$; ☹ = $\Delta i > 10$
Overall score: average from ☺ = 3 points; 😐 = 2 points; ☹ = 1 point.
AIC = Akaike information criterion; BIC = Bayesian information criterion; ERG = Evidence Review Group; KM = Kaplan-Meier

Table 5.9: Overall goodness-of-fit assessment by the ERG: busulfan EFS in the AML + MDS population

Parametric model	Visual fit with KM	AIC	BIC	Overall assessment [score]
Exponential	☹	☹	☹	☹ [1]
Weibull	☹	☹	☹	☹ [1]
Lognormal	☹	☹	☹	☹ [1]
Log-logistic	☹	☹	☹	☹ [1]
Gompertz	☺	☹	☺	☺ [2.7]
Gamma	☹	☺	☺	😐 [2.3]
MCM Weibull	😐	☹	☹	☹ [1.3]
MCM Lognormal	☺	☺	☺	☺ [3]
NMCM Weibull	😐	☹	☺	😐 [2]
NMCM Lognormal	☺	☺	☺	☺ [3]

Visual fit with KM: Based on company's assessment. The ERG is more neutral regarding this assessment.
AIC: Define $\Delta i = AICi - AICmin$, where AICi is the AIC of the i-th model, and AICmin is the lowest AIC, obtained for NMCM lognormal: ☺ = $\Delta i < 4$; 😐 = $4 < \Delta i < 7$; ☹ = $\Delta i > 10$.
BIC: Define $\Delta i = BICi - BICmin$, where BICi is the BIC of the i-th model, and BICmin is the lowest BIC, obtained for NMCM lognormal: ☺ = $\Delta i < 10$; ☹ = $\Delta i > 10$.
Overall score: average from ☺ = 3 points; 😐 = 2 points; ☹ = 1 point.
AIC = Akaike information criterion; BIC = Bayesian information criterion; ERG = Evidence Review Group; KM = Kaplan-Meier

Table 5.10: Overall goodness-of-fit assessment by the ERG: treosulfan OS in the AML + MDS population

Parametric model	Visual fit with KM	AIC	BIC	Overall assessment [score]
Exponential	☹	☹	☺	☹ [1.3]
Weibull	☹	😐	☺	😐 [2]
Lognormal	☹	☺	☺	☺ [2.7]
Log-logistic	☹	☺	☺	😐 [2.3]
Gompertz	☺	☺	☺	☺ [3]
Gamma	☺	☺	☺	☺ [3]
MCM Weibull	☺	☺	☺	☺ [3]
MCM Lognormal	☺	☺	☺	☺ [3]
NMCM Weibull	☺	☺	☺	☺ [3]
NMCM Lognormal	☺	☺	☺	☺ [3]

Visual fit with KM: Based on company's assessment. The ERG is more neutral regarding this assessment.
AIC: Define $\Delta i = AICi - AICmin$, where AICi is the AIC of the i-th model, and AICmin is the lowest AIC, obtained for Gompertz: ☺ = $\Delta i < 4$; 😐 = $4 < \Delta i < 7$; ☹ = $\Delta i > 10$.
BIC: Define $\Delta i = BICi - BICmin$, where BICi is the BIC of the i-th model, and BICmin is the lowest BIC, obtained for Gompertz: ☺ = $\Delta i < 10$; ☹ = $\Delta i > 10$.
Overall score: average from ☺ = 3 points; 😐 = 2 points; ☹ = 1 point.
AIC = Akaike information criterion; BIC = Bayesian information criterion; ERG = Evidence Review Group; KM = Kaplan-Meier

Table 5.11: Overall goodness-of-fit assessment by the ERG: busulfan OS in the AML + MDS population

Parametric model	Visual fit with KM	AIC	BIC	Overall assessment [score]
Exponential	☹	☹	☹	☹ [1]
Weibull	☹	☹	☹	☹ [1]
Lognormal	☹	☹	☹	☹ [1]
Log-logistic	☹	☹	☹	☹ [1]
Gompertz	☹	☹	☹	☹ [1]
Gamma	☹	😐	☺	😐 [2]
MCM Weibull	☺	😐	☺	☺ [2.7]
MCM Lognormal	😐	☺	☺	☺ [2.7]
NMCM Weibull	☺	😐	☺	☺ [2.7]
NMCM Lognormal	😐	☺	☺	☺ [2.7]

Visual fit with KM: Based on company's assessment. The ERG is more neutral regarding this assessment.
AIC: Define $\Delta i = AICi - AICmin$, where AICi is the AIC of the i-th model, and AICmin is the lowest AIC, obtained for NMCM lognormal: ☺ = $\Delta i < 4$; 😐 = $4 < \Delta i < 7$; ☹ = $\Delta i > 10$.
BIC: Define $\Delta i = BICi - BICmin$, where BICi is the BIC of the i-th model, and BICmin is the lowest BIC, obtained for NMCM lognormal: ☺ = $\Delta i < 10$; ☹ = $\Delta i > 10$.
Overall score: average from ☺ = 3 points; 😐 = 2 points; ☹ = 1 point.
AIC = Akaike information criterion; BIC = Bayesian information criterion; ERG = Evidence Review Group; KM = Kaplan-Meier

5.2.6.2 Long-term mortality

Two main aspects were considered by the company when modelling long-term mortality: the selection of a "cure point" and the application of a standardised mortality ratio (SMR) for background mortality.

The rationale for including a "cure point" in the model is that HSCT is a potentially curative treatment. Prior to this "cure point", transitions between the health states of the model are calculated based on the parametric OS and EFS curves. After the cure point, only background mortality is determining the transitions between the health states. This was model by the company by either using 1) life tables for the general UK population or 2) life tables with SMRs for HSCT applied.

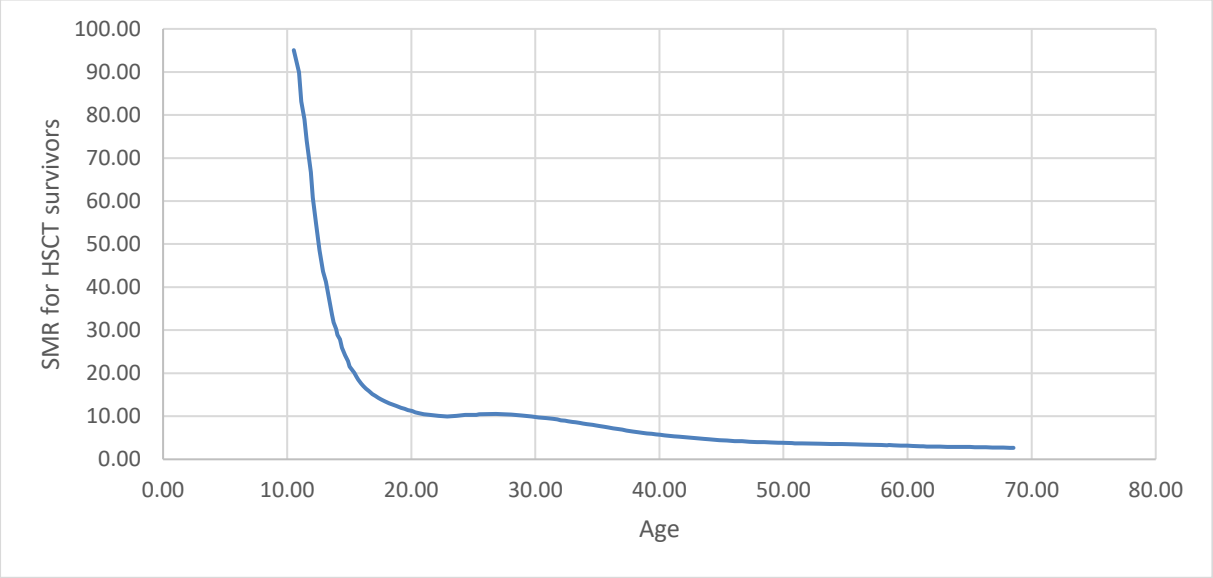
In the base-case, the company assumed a fixed "cure point" of five years. This choice was made based on the opinion of two clinical experts. In response to the clarification question C6, the company explained that patients surviving alloHSCT for at least five years are considered as "cured". This means that relapses/transplant-related deaths after five years are very rare. This is shown in the Kaplan-Meier estimates of OS from MC-FludT.14/L Trial II presented in Figure 2 in the response to the request for clarification (Figure 4.2).[12] This plot shows [redacted]. Furthermore, the company indicated that the impact of treosulfan on effectiveness is [redacted].[12]

The following SMRs for HSCT patients were included in the model:

- Assuming a mortality rate equal to the UK general population (SMR = 1)
- Assuming an overall SMR for all alloHSCT patients (SMR = 4.30). Sourced from supplementary Table A1 in Martin 2010.[49]
- Assuming an overall SMR for all patients older than 45 years (SMR = 3.20). Sourced from supplementary Table A3 in Martin 2010.[49]
- Assuming an SMR based on the mean age of the patients plus five years of the fixed "cure point" (SMR = 2.88). If this option is selected, the model uses an SMR calculated by adding the duration of the "cure point" onto the mean age of patients in the model cohort, and by selecting the closest SMR from a lookup table based on Figure 5.4 which was digitised from supplementary Figure A3C in Martin 2010[49]. For example, a cure point of five years in a cohort with a mean age of 60 years would use the SMR for patients aged 65 (the value for 65.03 in the digitised data, SMR = 2.88). The author deemed there to be too few data points available to generate meaningful estimates for patients younger than 10 and patients older than 70 years of age based on the spline-smoothed Poisson model that was used. In light of this, patients in the model aged above the end of the curve were assumed to have an SMR value equal to that at the end of the curve (SMR = 2.66).
- Hazard ratio (HR) used in the NICE TA552[41] (HR = 2.30). It was considered a plausible estimate by the ERG in TA552.
- Assuming an overall SMR for all HSCT patients (SMR = 4.50). Sourced from Table 3 in Martin 2010.[49]

All the above SMRs, except the first one, were sourced from Martin 2010,[49] a study reporting SMRs for a cohort of patients who underwent HSCTs for the treatment of several diseases including AML and MDS. The HR = 2.30 used in the TA552[41] was assumed for the base-case, since this was the value deemed as the most appropriate by the clinical experts consulted by the company.

Figure 5.4: SMRs for HSCT survivors



Based on Figure 16 of the CS[1]
CS = company submission; HSCT = haematopoietic stem cell transplant, SMR = standardised mortality ratio

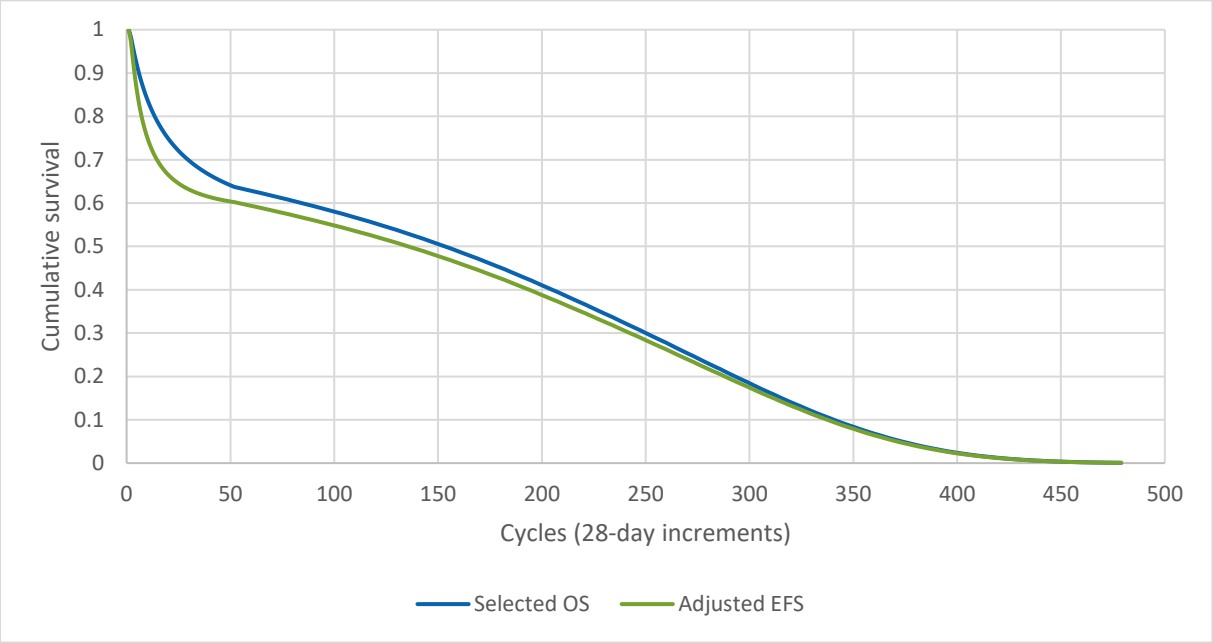
Besides the selection of the "cure point" and the SMR for background mortality, the company also included in the model the following five approaches to long-term mortality:

1. Use of (unadjusted) parametric curves extrapolations.
2. Use of parametric curves or general population survival based on life tables, depending on which has the highest mortality rate.
3. Use of parametric curves or HSCT-specific life tables (as determined by SMRs), depending on which has the highest mortality rate.
4. Use of parametric curves up to the fixed "cure point", followed by a switch to general population life table mortality rates.
5. Use of parametric curves up to the fixed "cure point", followed by switch to HSCT-specific life table (as determined by SMRs) mortality rates.

Approach 5 was assumed by the company for the base-case, since this was deemed to be the most appropriate and reflective of AML and MDS patients by the clinical experts consulted by the company. As pointed out by the company in the response to the clarification question C7,[12] the term "cure point" refers to patients being functionally cured. Therefore, in the model it is possible to relapse after the "cure point". Note the switch from parametric curves to SMR-adjusted life tables was applied to both OS and EFS curves, to allow a small proportion of patients to relapse after the "cure point". The underlying assumption, supported by the clinical experts consulted by the company, is that when patients have survived five years after HSCT, most of the mortality risk for this population will not be attributed to a relapse of AML and MDS, but to other causes, like long-term complications associated with HSCT itself.

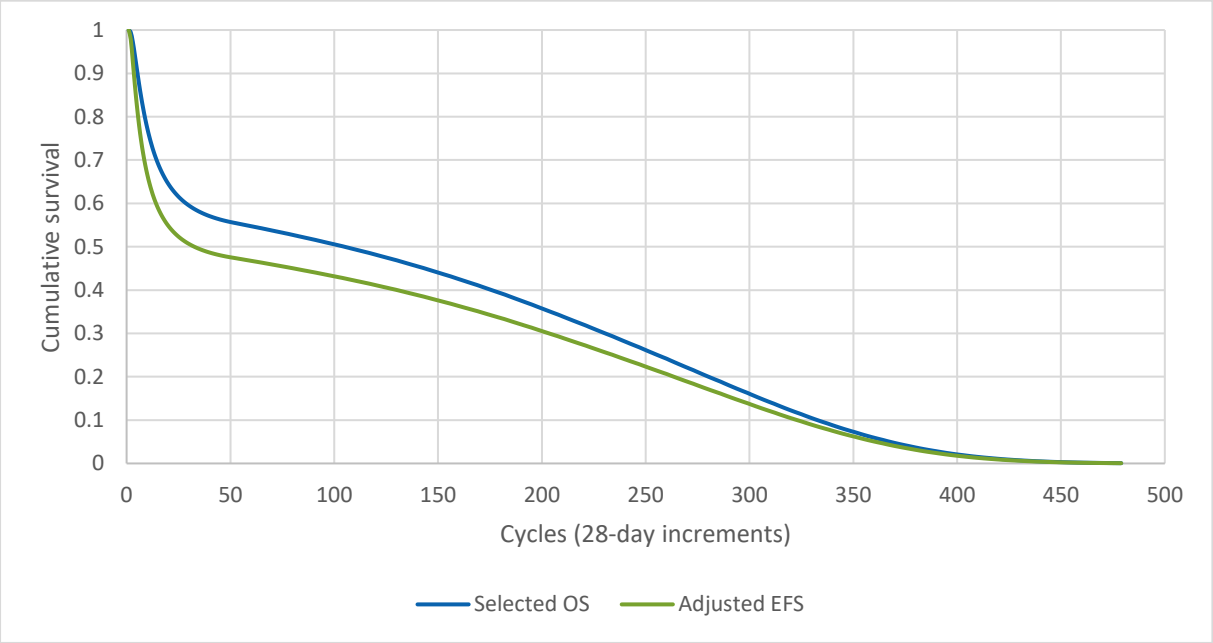
In summary, the company selected in the base-case analysis a lognormal NMCM distribution to model EFS and OS for both treosulfan and busulfan arms up to a five-year "cure point", followed by switch to an HSCT-specific life table mortality rates determined by an HR = 2.30. The resulting base-case OS and EFS curves for the pooled AML+MDS population for treosulfan and busulfan are shown below in Figures 5.5 and 5.6, respectively.

Figure 5.5: Base-case OS and EFS curves for the pooled AML + MDS population, treosulfan arm



Based on Figure 17 of the CS[1]
AML = acute myeloid leukaemia, CS = company submission; EFS = event-free survival, MDS = myelodysplastic syndrome, OS = overall survival

Figure 5.6: Base-case OS and EFS curves for the pooled AML + MDS population, busulfan arm



Based on Figure 18 of the CS[1]
AML = acute myeloid leukaemia, CS = company submission; EFS = event-free survival, MDS = myelodysplastic syndrome, OS = overall survival

ERG comment: The selection of the five-year "cure point" was made following the feedback from two clinical experts consulted by the company. To account for the uncertainty around this value, the company's electronic model allows selecting values smaller than five years. However, the company did

not present any economic results based on different "cure points". The impact of choosing different "cure points" on the economic results is explored by the ERG in Section 7 of this report.

The two clinical experts consulted by the company also suggested that the HR = 2.3 from TA552[41] was the most plausible estimate for calculating long-term mortality of patients post-HSCT. However, the impact of choosing different SMRs on the economic results was not assessed in the CS. The ERG explored this with additional scenarios in Section 7 of this report.

The company presented in the CS five different approaches to model long-term mortality. In the response to the clarification question C8,[12] the company confirmed that only approaches 3 (use of parametric curves or HSCT-specific life tables – as determined by SMRs – depending on which has the highest mortality rate) and 5 (use of parametric curves up to the fixed "cure point", followed by switch to HSCT-specific life table – as determined by SMRs – mortality rates) are plausible. The company also indicated that approaches 1, 2 and 4 were considered as potential options required to derive the plausible scenarios (3 and 5). These were included in the model for transparency and it was useful for internal validation. The ERG agrees with this statement but would like to note that the impact of selecting different approaches to long-term mortality was also not explored by the company in the CS. In Section 7 of this report, the ERG explores this but focused only on the plausible approaches (3 and 5).

Regarding approach 3 to long-term mortality (use of parametric curves or HSCT-specific life tables – as determined by SMRs – depending on which has the highest mortality rate), the ERG found the definition of the method unclear and asked the company to clarify several aspects of this method in clarification question C5.[12] In their response, the company confirmed that, unlike approach 5, there is no fixed "cure point", since this is determined by the time point where the HSCT-specific mortality rate is highest. Thus, this approach overrides the cure point assumption of five years.

The ERG also requested the company to provide estimates of the "cure points" obtained using method 3 and to discuss their clinical validity. The cure points were different for each OS and EFS model arm and they are summarised in Table 5.12. Finally, the company mentioned that, for this reason, this approach was not selected for their base-case analysis. The company discussed the clinical validity of these "cure points" with experts, who suggested that having a fixed "cure point" of five years would be the most appropriate assumption.

Table 5.12: Cure points when the third approach to survival modelling (choosing the highest rate of mortality using either the parametric models or HSCT-adjusted life tables)

Survival curve	Model arm	Approximate cure point, years
OS	Treosulfan: pooled cohort	6.8
	Treosulfan: AML only	6.6
	Treosulfan: MDS only	6.8
	Busulfan: pooled cohort	5.2
	Busulfan: AML only	4.4†
	Busulfan: MDS only	5.5
EFS	Treosulfan: pooled cohort	4.6
	Treosulfan: AML only	4.2
	Treosulfan: MDS only	5.8

Survival curve	Model arm	Approximate cure point, years
	Busulfan: pooled cohort	5.1
	Busulfan: AML only	5.2
	Busulfan: MDS only	4.8

Based on Table 1 in response to the request for clarification.[12]
† This estimate excludes the first model cycle, where the HSCT-adjusted life table mortality estimate is instantaneously higher than the mortality estimate from the parametric curve for the first 28-day period.
AML = acute myeloid leukaemia; EFS = event-free survival; MDS = myelodysplastic syndrome; OS = overall survival

Finally, as requested by the ERG, the company amended approach 5 to long-term mortality to allow the possibility to be considered functionally cured earlier than the five-year cure point. In theory, this occurred in the new version of the model when the mortality rate on the EFS parametric curve exceeded that of the SMR-adjusted HSCT mortality. However, this amendment did not change the results at all since none of the EFS mortality rates were lower than those with the HSCT-adjusted life tables before five years. In order to see any difference in the results, patients would need to be considered cured earlier than four years.

5.2.7 Adverse events

Since it is a well-known complication of HSCT, Graft- versus-host disease (GvHD) was taken into account as a relevant adverse event (AE) in the cost effectiveness model. Both stage III/IV acute GvHD (aGvHD; i.e. occurring in the first 100 days of the trial) and chronic GvHD (cGvHD; i.e. occurring between days 100 and 731), with the incidences based on findings from MC-FludT.14/L Trial II. These incidences are displayed in Table 5.13.

In addition to extensive cGvHD and stage III/IV aGvHD, all treatment related Grade 3+ Common Terminology Criteria for Adverse Events (CTCAEs) with an incidence ≥1% in MC-FludT.14/L Trial II were included in the model (see Table 5.13). As cGvHD cannot occur until day 100, incidence estimates reflect events between day 100 and day 731 in the trial, with aGvHD events for the first 100 days, and all other AEs for the first 28 days.

Table 5.13: Incidence of treatment related adverse events with ≥1% incidence from the phase III clinical trial

Adverse event	Cumulative incidence	
	Busulfan	Treosulfan
Extensive cGvHD (days 100-731)	26.7%	19.7%
Stage III/IV aGvHD (up to 100 days)	8.1%	6.4%
Mucositis oral	6.0%	4.4%
Nausea	4.9%	2.6%
Diarrhoea	0.7%	1.1%
Vomiting	1.4%	0.4%
Gamma glutamyl transferase (GGT) increased	8.1%	3.0%
Alanine aminotransferase increased	2.8%	4.8%
Aspartate aminotransferase increased	2.1%	4.1%

Adverse event	Cumulative incidence	
	Busulfan	Treosulfan
Blood bilirubin increased	1.4%	2.2%
Investigations (other)	1.4%	1.5%
Febrile neutropenia	4.9%	4.4%
Sepsis	0.4%	2.2%
Lung infection	0.7%	1.1%
Anorexia	1.4%	1.5%
Syncope	1.4%	0.0%
Rash maculo-papular	1.1%	0.7%

Based on Table 9 of Appendix I in the CS.[50]
aGvHD = acute graft vs host disease; cGvHD = chronic graft vs host disease; CS = company submission; GGT = gamma glutamyl transferase.

For extensive cGvHD and Stage III/IV acute GvHD, mean duration estimates of nine months and 2.5 months were used respectively, based on TA545,[44] and validated through clinical expert opinion. Other AEs were assumed to have an event duration equal to the average inpatient stay for the associated HRG codes, sourced via NHS reference cost data.[51] Due to lack of appropriate codes in the latest NHS reference costs, NHS Tariff data were used for the costing of the following AEs: gamma glutamyl transferase (GGT) increased, alanine aminotransferase increased, aspartate aminotransferase increased, blood bilirubin increased, investigations (other), and febrile neutropenia. The durations of the AEs that were costed using NHS Tariff data were based on findings from MC-FludT.14/L Trial II. An overview of the durations of AEs is provided in Table 5.14.

Table 5.14: Overview of the durations of AEs

Adverse event	Duration, days	
	Mean (SE)	Source
Extensive cGvHD	273.9 (54.8)*	Assumption based on TA545[44] and clinical expert opinion
Stage III/IV aGvHD	76.1 (15.2)*	Assumption based on TA545[44] and clinical expert opinion
Mucositis oral	2.2 (0.4)*	NHS reference costs 2017/18[51] Weighted average of elective inpatient currency codes CB02A-CB02F (Non-Malignant, Ear, Nose, Mouth, Throat or Neck Disorders, with Interventions).
Nausea	3.3 (0.7)*	NHS reference costs 2017/18[51] Weighted average of elective inpatient currency codes FD10A-FD10M (Non-Malignant Gastrointestinal Tract Disorders).
Diarrhoea	3.3 (0.7)*	NHS reference costs 2017/18[51] Weighted average of elective inpatient currency codes FD10A-FD10M (Non-Malignant Gastrointestinal Tract Disorders).

Adverse event	Duration, days	
	Mean (SE)	Source
Vomiting	3.3 (0.7)*	NHS reference costs 2017/18[51] Weighted average of elective inpatient currency codes FD10A-FD10M (Non-Malignant Gastrointestinal Tract Disorders).
GGT increased	40 (4.7)	Analysis of phase III trial data
Alanine aminotransferase increased	15.1 (4.2)	Analysis of phase III trial data
Aspartate aminotransferase increased	8.7 (1.8)	Analysis of phase III trial data
Blood bilirubin increased	23.4 (8.1)	Analysis of phase III trial data
Investigations (other)	31.3 (11.8)	Analysis of phase III trial data
Febrile neutropenia	4.1 (1.1)	Analysis of phase III trial data
Sepsis	6.1 (1.2)*	NHS reference costs 2017/18[51] Weighted average of elective inpatient currency codes WJ06A-WJ06J (Sepsis).
Lung infection	5.4 (1.1)*	NHS reference costs 2017/18[51] Weighted average of elective inpatient currency codes DZ11K-N;P-V (Lobar, Atypical or Viral Pneumonia).
Anorexia	3.2 (0.6)*	NHS reference costs 2017/18[51] Weighted average of elective inpatient currency codes FD04A-E (Nutritional Disorders).
Syncope	2 (0.4)*	NHS reference costs 2017/18[51] Weighted average of elective inpatient currency codes EB08A-E (Syncope or Collapse).
Rash maculopapular	16 (6)	Analysis of phase III trial data

Based on Table 10 in Appendix I of the CS.[50]
* SE assumed to be 20% of the mean
aGvHD = acute graft vs host disease; BNF = British National Formulary; cGvHD = chronic graft vs host disease; CS = company submission; GGT = gamma glutamyl transferase; HCHS = Hospital and Community Health Services; NHS = National Health Service; SE = standard error

Mean duration estimates were used to spread AE disutilities and costs over multiple cycles, assuming a constant rate of incidence over time. Incidence start cycle and incidence end cycle weights were estimated to partition incidence estimates for GvHD events that were only possible to occur for a particular time frame within a 28-day cycle.

ERG comment: The company has included a number of relevant AEs that aligns with the AEs with an incidence ≥ 1% based on data from MC-FludT.14/L Trial II. The ERG notes that as such, the following complications of HSCT that were mentioned in Section B.1.3.5.1 in the CS are not included as AEs in the model: prolonged, severe pancytopenia; graft rejection, hepatic sinusoidal obstruction syndrome, and late onset problems due to HSCT.[1] Furthermore, it is not clear what specific AE is being referred

to as 'Investigations (other)'. However, the ERG notes that the impact of AEs on the cost effectiveness results is small. Overall, the ERG considers the set of AEs that have been incorporated as appropriate.

5.2.8 Health-related quality of life

5.2.8.1 Identification and selection of utility values

Health-related quality of life data was not collected in the trial. The company searched for published sources of health state utility values through a systematic literature review, a review of prior technology appraisals involving AML and MDS populations and additional targeted searching of the literature. Two economic modelling papers were identified in the systematic review, of which one contained health state utility values estimated using visual analogue scale.[52] However, this study was excluded as only median values were provided.

Thirty-two studies providing data on HRQoL of AML, MDS, ALL and MM patients were identified in the systematic review. Fifteen of these studies provided utility values derived using preference based utility measures, while the remaining 17 utilised non-preference based measures of QoL. Eighteen studies (six preference based, 12 non-preference based) were excluded as they provided data only for ALL or MM patients, who were not covered in the trial. Two further preference based studies were excluded as they did not provide fully appropriate utilities for the health states in the model.[53, 54] Three non-preference based studies were excluded as they did not provide sufficiently detailed quality of life results to enable mapping for the required health states in the model.[55-57] The remaining non-preference based study, which considered AML patients in long-term remission, was considered for inclusion.[58] However, it was subsequently excluded due to face validity concerns surrounding the QLQ-C30 domain scores and mapped EQ-5D estimates for long-term remission patients and wide variety of mapped estimates that were obtained using different mapping algorithms. The remaining seven preference based studies were considered appropriate for inclusion in the model.[47, 59-64] However, none of these contained UK EQ-5D data as preferred by NICE.

In response to the lack of utility data matching the NICE reference case, the company also searched for HRQoL data in relevant technology appraisals and conducted an additional targeted review of the literature. From the targeted review, four additional sources of non-UK EQ-5D utility values,[48, 65-67] three sources of non-EQ-5D preference based values,[1, 68-70] and four studies including non-preference based HRQoL data were identified.[71-74] Additionally, two sources of mapped QLQ-C30 to EQ-5D values (using algorithms by Proskorovsky et al. 2014 and McKenzie et al. 2009[75, 76]) were identified from TA399[43], which were used as base-case values in TA545.[44] One study was excluded as it only reported values of another included study.[69] Another study was excluded due to very small sample sizes for the health states required by the model.[66]

Quality of life data from Grulke et al. 2012 was used to estimate post-HSCT recovery health state utility values in the base-case.[71] This study investigated the quality of life of patients before and after haematopoietic stem cell transplantation, by gathering data from 38 samples from 33 papers in English and German that provided data using the EORTC Quality of Life Core Questionnaire QLQ-C30. This study was chosen for the base-case because the health state utility values included were estimated from much larger samples than other studies identified and the study provided five time-based health state utility values for patients post-HSCT, which allowed for patients quality of life post-HSCT to vary as time from procedure increased. The values obtained from mapping Grulke were also judged by the company to have better face validity than those produced by Kurosawa et al. 2016[63], which also provided multiple time-based post-HCST utility values. Values obtained from TA399[43] and Uyl de Groot were also excluded from the base-case as they provided health state values which were higher

than age matched UK EQ-5D general population norms.[43, 67] The Grulke utilities were adjusted for age and gender using methods from Ara and Brazier 2010[77] to compensate for the fact that the population in the Grulke study were considerably younger (40-45 years old) than the trial population.[71]

ERG comment: The ERG considers that the search for utility values was comprehensive and the reasons provided for excluding or selecting utility sources were generally well reported and justified. The ERG agree that, among the relatively large number of potential sources, the utility values from Grulke et al. 2012[71] are a good option for the base-case, given the large sample size and availability of post-HSCT time dependent utility values. The model also provides the option to use a range of appropriate utility values from alternative sources, in order to test the impact of the choice of utility values on the ICER. However, the choice of utility values does not appear to have a large impact.

5.2.8.2 Identification and selection of mapping algorithms

As no UK EQ-5D values were identified, the company also identified mapping algorithms to enable them to convert non-preference based HRQoL data into EQ-5D utility values. Only one relevant mapping algorithm was identified from the systematic review.[75] The company searched for additional algorithms using the Health Economics Research Centre (HERC) mapping database,[78] prior technology appraisals in AML/MDS and targeted review of the literature. Mapping studies were excluded if they did not use UK EQ-5D tariffs in the study, or if the models required patient level data. The choice of which algorithm to use was based primarily on the patient population characteristics of the original mapping study including: indication, UK patient representation, number of observations in estimation sample as well as predictive performance ranking criteria (mean absolute error (MAE), root mean square error (RMSE), MAE on specific EQ-5D-3L score subsets of 0.75 - 1, 0.5-0.75, 0-0.5 and <0 and error in the mean predicted QALYs compared to the mean observed QALYs) as recommended by Longworth et al. 2014.[79]

Six mapping algorithms were identified for the QLQ-C30, with one subsequently excluded due to a lack of reporting of intercept coefficients for the linear mixed model.[80] The characteristics of the remaining five algorithms are summarised in Table 42 of the company submission.[1] Proskorovsky et al. 2014[75] was selected for the base-case analysis as it included a large proportion of UK patients (58%) and the multiple myeloma patient group was considered most similar to the model population, particularly as 12% of the sample had received prior autologous stem cell transplant. Moreover, this algorithm had been found to perform well in terms of MAE, particularly in EQ-5D values>0.5 (where most patients in the long-term model would be represented), in a mapping validation study of a large set of cancer patients by Doble et al. 2016.[81] In addition, the company judged the Proskorovsky et al. 2014[75] algorithm to produce the most plausible set of mapped EQ-5D values for Grulke et al. 2012[71], which was selected for the base-case analysis to estimate post-HSCT recovery/remission utilities. These values were considered most plausible as they resulted in continuous increases in utilities for each time point post-HSCT and provided year 4+ post-HSCT utility estimates that were considered to be the most plausible reflection of disutility for functionally cured patients, compared to the general population (see Section B.3.3.2.5). Furthermore, Proskorovsky et al. 2014[75] was used in the base-case analysis for TA399,[43] which was identified as a key source of relapse utilities and adverse event disutilities in the company base-case. As such, application of the Proskorovsky algorithm provided consistency in application of mapping algorithms used across health states in the model. The McKenzie et al. 2009[76] algorithm was selected for scenario analysis as it had the largest total number of observations in the estimation sample (877), an exclusively UK population, the most conservative set of post-HSCT recovery/remission utilities, allowed for internal consistency with use of mapped EQ-5D estimates from

TA399[43] and performed better than the other algorithms in a multiple myeloma mapping validation study identified in the systematic review.[82]

Six studies were identified which mapped from the SF-36 to EQ-5D. One was excluded as it used the Korean EQ-5D tariff[83] another due to inappropriate population (frozen shoulder)[83] and two as the company could not find the original paper.[84, 85] This left two included mapping studies, summarised in Table 44 of the company submission.[1] Both algorithms produced plausible EQ-5D values as compared to the general population for the two cGvHD quality of life studies with SF-36 data considered in the economic analysis.[72, 73] While the SF-36 mapping algorithms were not applied in the base-case analysis, the Rowen 2009[86] algorithm was preferred by the company due to the larger observation estimation sample size, lower RMSE and more appropriate patient population (hospital inpatients and outpatients) for alloHSCT patients with AML and MDS.

Eight studies were identified in the HERC database[78] for mapping SF-12 to EQ-5D. The only study with SF-12 data utilised in the model,[74] which studied the quality of life of cGvHD and aGvHD/cGvHD overlap patients, only provided average physical and mental composite scores. Therefore, this was a primary criterion for selecting the most appropriate algorithm for the SF-12, alongside the use of the UK tariff, population and provision of sufficient information to carry out mapping. Five studies were excluded based on these criteria, leaving the two studies summarised in Table 46 of the company submission.[1] Mapped SF-12 estimates were not required in the base-case analysis. However, the Franks algorithm[87] was preferred over Lawrence[88] due to a larger number of estimation sample observations and provision of standard error estimates for the coefficients. Both mapping studies were included in order to allow for potential exploration of disutility specific to aGvHD (applied as a disutility vs cGvHD).

ERG comment: Again, the search for mapping algorithms was generally comprehensive and well reported. The ERG concurred with the use of the Proskorovsky et al. 2014[75] mapping algorithm in the base-case, alongside the Grulke et al. 2012[71] health state utility values, as the population on which the Proskorovsky algorithm was developed was most similar to the population in the current appraisal, both in terms of condition and geography. Again, the company allows for alternative algorithms to be tested in the model to allow the impact of this choice to be tested.

5.2.8.3 Adverse event disutilities

Several studies were identified in the literature review that included quality of life data for GvHD.[48, 59, 63, 65, 72-74] However, none of these studies met the NICE reference case requirements and required mapping. The Kurosawa studies contained EQ-5D data, but based on non-UK valuation tariffs.[48, 63, 65] Two studies contained SF-36 data,[72, 73] which were mapped using the Rowen and Ara algorithms described in Section 5.2.8.2.[86] One study was mapped from SF-12 using Franks and Lawrence algorithms,[74] and one study was based on DCE/TTO methods.[59]

None of the identified studies provided health state quality of life data specific to extensive chronic GvHD (cGvHD) or stage III/IV acute GvHD (aGvHD). Two studies were identified which provided SF-36 data based on other classifications of cGvHD severity.[72, 73] The Peric study used definitions of "inactive" and "active/highly active" within the cGvHD group while the Lee study used classifications of mild/moderate/severe cGvHD.[72, 73] Clinical experts were unfamiliar with the definitions in the Peric study so these data were not used. On advice from the clinical expert, patients with moderate to severe cGvHD would likely be similar to those with extensive cGvHD. However, these values were not used in the base-case or scenario analyses as there was no evidence on how to appropriately weight the

moderate and severe cGvHD utilities to reflect this population and "no cGvHD" reference group to estimate the disutility.

Only one study identified provided data on the quality of life of aGvHD patients.[74] However this data only included values for patients with overlapping aGvHD and cGvHD symptoms, which were not necessarily appropriate to apply to aGvHD patients only in the model. There were also issues with the internal consistency of the mapped EQ-5D estimates of patients with cGvHD only compared to patients with no cGvHD at six months post-HSCT. Therefore, an aGvHD specific disutility was not included in the model. The same disutility for both cGvHD and aGvHD was assumed in the model. This assumption was considered reasonable by clinical experts.

In the base-case a GvHD disutility of 0.120 was used, based on Kurosawa et al. 2016.[63] This source was considered to most closely align with the NICE reference case and to be more appropriate than applying mapped estimates from SF-36 and SF-12 using mapping algorithms developed in non-cancer populations. A GvHD disutility of 0.190 from Castejon et al. 2018[59] was applied in the TTO/DCE utility scenario analysis.

Very few published studies were identified investigating the impact of other Grade 3+ CTCAEs on the quality of life of AML and MDS patients receiving alloHSCT. One study by Stein et al. 2018[62] was identified which provided disutility estimates for other AEs relevant to the model (serious infections, severe diarrhoea, severe redness/skin peeling and abnormally low blood counts). However, these disutilities were generated using DCE. Two sets of disutilities for Grade 3 and 4 AEs were identified from TA399[43] (0.0240 and 0.0207) based on different mappings of the QLQ-C30 data from the clinical trial of azacitidine, a chemotherapy based treatment. Despite concerns regarding differences between AEs experienced by patients on chemotherapy treatment vs HSCT, these values were judged to be more aligned with the NICE reference case than values estimated using DCE. Therefore, these disutilities were applied in the base-case model for all other adverse events, with values from Stein et al. 2018[62] explored in scenario analysis.

ERG comment: The ERG agrees with the company's choice to use the disutilities for Grade 3 and 4 AEs identified from TA399[43] (0.0240 and 0.0207), despite concerns surrounding differences between AEs experienced by patients on a chemotherapy based treatment versus patients receiving alloHSCT. The larger disutility values from Stein et al. 2018[62] will be explored in scenario analysis but, as the company stated in the clarification response, when asked to consider the impact of this choice of base-case on the model results, the impact is likely to be small.[12]

5.2.8.4 HRQoL data used in the cost effectiveness analysis

Upon entering the model, all patients begin in the induction/HSCT health state. The HSCT induction utility is applied to all patients for one cycle. After the first cycle, patients move to the HSCT recovery (remission) state, where they remain until they experience a relapse event. The HSCT state is split into separate time based states of ≤6 months, 7-12 months and years 2, 3 and 4+. Separate time based utilities are utilised to reflect increasing improvement in quality of life following HSCT. For patients who experience relapse, utilities for AML and MDS relapse are applied. Due to a lack of data for MDS specific relapse, the same relapse utility is used for both condition groups in the base-case. However, several MDS specific relapse utility values are available for scenario analysis.

All health state utility values used in the model base-case are adjusted for age and sex, using the equation provided in Ara and Brazier 2010.[77] All health state utility values used in the model base-case are calculated by creating a utility multiplier for being in a certain state. This multiplier is calculated by

dividing the utility value sourced for that health state by the age and gender adjusted general population utility value (age adjusted according to the mean age of the source study) calculated from Ara and Brazier 2010.[77] This multiplier is then multiplied by the relevant age and gender adjusted general population utility value, according to the age of patients currently in that health state in the model.

Table 5.15 shows the base-case utility values (unadjusted for age and gender) used in the model for all health states in the model.

Table 5.15: Health state utility values for base-case

Health state	Base-case Value (unadjusted for age or gender)	Source
Induction/HSCT	0.558	Grulke et al. 2012[71] Proskorovsky et al. 2014[75]
Post HSCT recovery (short-term)		
Discharge	0.660	
≤6 months	0.756	
7-12 months	0.818	
Post HSCT recovery (long-term)		
Year 2	0.822	
Year 3	0.822	
Year 4+	0.870	
Functionally cured	0.870	(set equal to Year 4+)
Relapse/progression		
AML	0.623	TA399[43] / TA545[44] Proskorovsky et al. 2014[75]
MDS	0.623	(set equal to AML)
GvHD disutilities		
Extensive chronic GvHD	-0.12	Kurosawa et al. 2016[63]
Stage III-IV acute GvHD	-0.12	(set equal to cGvHD)
Other adverse event disutilities		
Grade III+ AEs	-0.024	TA399[43] / TA545[44] Proskorovsky et al. 2014[75]
Serious infection	-0.024	(set equal to Grade III+ AEs)
Serious diarrhoea	-0.024	
Severe redness/skin peeling	-0.024	
Abnormally low blood counts	-0.024	

Based on Table 51 of the CS[1] (CS reports gender and age-adjusted utility values – this table reports unadjusted)
AE = adverse event; AML = acute myeloid leukaemia; cGvHD = chronic graft-versus-host disease; CS = company submission; GvHD = graft-versus-host disease; HSCT = Haematopoietic Stem Cell Transplantation; MDS = myelodysplastic syndrome; TA = technology appraisal

ERG comment: The ERG found no issues in the implementation of the utilities in the model. Choices of base-case utility values will be explored in sensitivity analyses, as shown in Table 5.16.

Table 5.16: Health state utility values for ERG scenario analyses

Health state	Scenario Value (unadjusted for age or gender)	Source
Utility scenario 1 – McKenzie mapping on Grulke values (all other values equal to base-case)		
Induction / HSCT utility	0.361	Grulke et al. 2012[71] McKenzie et al. 2009[76]
Post-HSCT recovery (short term) discharge	0.740	
Post-HSCT recovery (short term) ≤6 months	0.740	
Post-HSCT recovery (short term) 7-12 months	0.740	
Post-HSCT recovery (long-term) year 2 and year 3	0.707	
Post-HSCT recovery (long-term) year 4+	0.746	
Functionally cured	0.746	
Relapse / progression (AML and MDS)	0.568	TA399[43] / TA545[44] McKenzie et al. 2009[76]
Utility scenario 2 – DCE values		
Induction / HSCT utility	0.280	Castejon et al. 2018[59]
Post-HSCT recovery (short term) discharge	0.620	
Post-HSCT recovery (short term) ≤6 months	0.620	
Post-HSCT recovery (short term) 7-12 months	0.620	
Post-HSCT recovery (long-term) year 2 and year 3	0.620	
Post-HSCT recovery (long-term) year 4+	0.620	
Functionally cured	0.760	
Relapse / progression (AML and MDS)	0.100	
Extensive chronic GvHD and stage III/IV acute GvHD disutility	-0.190	
Sepsis/lung infection	-0.218	Stein et al. 2018[62]
Diarrhoea	-0.176	
Oral mucositis and maculopapular rash	-0.060	
Febrile neutropenia	-0.100	
Utility scenario 3 - MDS specific relapse utility (all other values equal to base-case)		
Relapse / progression (AML)	0.623	TA399[43] / TA545[44] Proskorovsky et al. 2014[75]

Health state	Scenario Value (unadjusted for age or gender)	Source
Relapse / progression (MDS)	0.650	Szende et al. 2009[64]
Utility Scenario 4 – TA399/TA545 Proskorovsky values (all other values equal to base-case)		
Induction / HSCT utility	0.716	TA399[43] / TA545[44] Proskorovsky et al. 2014[75]
Post-HSCT recovery (short term) discharge	0.771	
Post-HSCT recovery (short term) ≤6 months	0.771	
Post-HSCT recovery (short term) 7-12 months	0.771	
Post-HSCT recovery (long-term) year 2 and year 3	0.771	
Post-HSCT recovery (long-term) year 4+	0.771	
Functionally cured	0.771	
Relapse / progression (AML and MDS)	0.623	
Extensive chronic GvHD and Stage III-IV acute GvHD	-0.120	Kurosawa et al. 2016[63]
Grade III+ AEs	-0.024	TA399[43] / TA545[44] Proskorovsky et al. 2014[75]

Based on Table 51 of the CS [1] (CS reports sex and age-adjusted utility values – this table reports unadjusted)
AE = adverse event; AML = acute myeloid leukaemia; cGvHD = chronic graft-versus-host disease; CS = company submission; GvHD = graft-versus-host disease; HSCT = Haematopoietic Stem Cell Transplantation; MDS = myelodysplastic syndrome; TA = technology appraisal

5.2.9 Resources and costs

5.2.9.1 Intervention costs and resource use

Following the regimens used in MC-FludT.14/L Trial II, treatment costs were applied for treosulfan and busulfan (both administered IV). Concomitant treatments used in busulfan patients were phenytoin, fludarabine, anti-thymocyte globulin (ATG; MUD patients only), ciclosporin (i.v. and oral), methotrexate and calcium folinate. Excluding phenytoin, the same therapies were used concomitantly with treosulfan patients. Administration costs were excluded from the model, as there was no difference in the administration time for treosulfan and busulfan in the clinical trial (120 minutes), and it was assumed that they were captured as part of the inpatient costs estimated for the HSCT procedure.

Treatment costs were sourced from the British National Formulary (BNF). For all concomitant treatments, dosing was based on the least costly pack costs per mg, with no wastage costs applied.

HSCT procedure costs were estimated from NHS reference costs, using Healthcare Resource Group (HRG) codes for bone marrow and peripheral cell harvest costs, and HSCT procedure codes for adult patients. A weighted average of the harvesting and HSCT procedure costs was used to calculate an average HSCT cost for each treatment arm, with the average overall cost for treosulfan and busulfan applied for secondary HSCTs.

For treosulfan and busulfan, wastage costs were applied, with 100% vial wastage assumed in the base-case model based on clinical expert opinion.

Treatment costs and HSCT procedure costs are summarised in Table 5.17.

Table 5.17: Treatment costs and HSCT procedure costs

Concomitant therapy	Method of administration	Dosing regimen	Unit cost (per vial/tablet)	Cost per mg	Total treatment cost (including wastage)	Notes/Comments
Treosulfan	i.v.	10 g/m^2 at days -4 to -2 prior to alloHSCT (within 120 minute period)	£53.83 (1g vial) and £208.03 (5g vial)	£0.05 (1g vial) and £0.04 (5g vial)	£2,496.41	Total cost excluding wastage: £2,410.28
Busulfan	i.v.	4 x 0.8 mg/kg/day at days -4 and -3 prior to alloHSCT (within 120 minute period)	£191.19	£3.19	£3,059.00	Total cost excluding wastage: £1,635.55
Fludarabine	i.v.	30 mg/m2 at days -6 to -2 prior to alloHSCT (within 30 minute period)	£147.07	£2.94	£851.96	Based on price for Fludara 50 mg powder for solution for injection vials
ATG	i.v.	10 mg/kg at days -4 to -2 prior to alloHSCT (Grafalon) or 2.5 mg/kg at days -2 and -1 prior to alloHSCT (Thymoglobuline)	£158.77	£6.35	£1,945.82	No BNF price available for Grafalon – treatment costs/dosing based on Thymoglubuline Weighted based on proportion of alloHSCTs with MUDs (76.4%)
Ciclosporin	i.v.	3 mg/kg/day i.v. at start (day -1 before and day of alloHSCT)	£11.01	£0.04	£21.18	Based on price for Sandimmun 250 mg/5 ml concentrate for solution for infusion ampoules
Ciclosporin	Oral	5 mg/kg/day oral (days +1, +3 and +6 after alloHSCT)	£2.28	£0.02	£27.38	Based on 100 mg capsule Drug Tariff (Part VIIIA Category C) price
Methotrexate	i.v.	15 mg/m2/day (dose 1 at day +1 after alloHSCT) then 10 mg/m2/day (doses 2 and 3 at days +3 and +6 after alloHSCT)	£200.57	£0.04	£2.71	Based on price for 5 g/200 ml solution for infusion vials

Concomitant therapy	Method of administration	Dosing regimen	Unit cost (per vial/tablet)	Cost per mg	Total treatment cost (including wastage)	Notes/Comments
CA-Folinate	i.v.	15 mg/m2 (dose 1 at day +1 after alloHSCT) then 10 mg/m2/day (doses 2 and 3 at days +3 and +6 after alloHSCT)	£4.62	£0.15	£10.42	Based on price for Refolinon 30mg/10ml solution for injection ampoules
Phenytoin	Oral	3 x 200 mg at day -5 before alloHSCT, then 3 x 100 mg at days -4 to -2 prior to alloHSCT	£0.33	£0.001	£3.14	Based on 300 mg capsule Drug Tariff (Part VIIIA Category A) price
HSCT procedure	-	-	-	-	£40,774.35	Calculated as the sum of the weighted average of harvesting costs (HRG codes SA18Z and SA34Z) and allogeneic transplant costs (HRG codes SA20A-SA23A, SA38A, SA39A, SA40Z) for elective inpatients
Total treosulfan					**£46,130.24**	
Total busulfan					**£46,695.97**	

Based on Table 10 from Appendix I of the CS.[50]
alloHSCT = allogeneic haematopoietic stem cell transplant; ATG = anti-thymocyte globulin; CS = company submission; i.v. = intravenous, MUD = matched unrelated donor

ERG comment: Since the costs of treosulfan and busulfan, and the costs of concomitant medication, were based on the regimens received by patients in MC-FludT.14/L Trial II, the ERG considers these as appropriate estimates. The high costs due to wastage of busulfan are caused by a combination of 1) the assumption of 100% vial wastage (based on expert opinion), 2) the average weight of the trial patients (indicating a dosage that is just slightly over the available vial size), and 3) the only available vial size (identified by the BNF) being 60 mg. This results in nearly half of the total amount of busulfan being wasted. The underlying rationale for the wastage was confirmed by the company in the response to the ERG's clarification questions. Following an updated summary of the product characteristics, a scenario (scenario 5 in the CS) was run where busulfan dosage was based on a single daily dose. This scenario resulted in substantially reduced wastage (costs). However, this did not affect the overall conclusions of the economic analysis. The costs of an HSCT procedure were, according to the ERG, appropriately based on a weighted average of the costs for harvesting of stem cells, from either bone marrow or peripheral blood, and the costs of various HSCT procedures for adult patients.

5.2.9.2 Health state unit costs and resource use

Event-free survival

Health state costs for event-free survival were sourced from an NHS Blood and Transplant Analysis 2015.[89] The costs associated with HSCT after the procedure itself were inflated to 2017/2018 values using the Health Services (HS) index published by the Personal Social Services Research Unit (PSSRU). Based on key opinion leader (KOL) input, after the initial two-year period the costs of HSCT patients who had not relapsed are those of three times monitoring per year, consisting of an outpatient haematology appointment, blood biochemistry, full blood counts and phlebotomy. The costs associated with HSCT after the procedure itself are listed in Table 5.18.

Table 5.18: Post-HSCT recovery costs (EFS)

Time post-HSCT	2017/2018 inflated costs	Source
0 to 6 months post-HSCT (cost per cycle)	£4,959.47	NHS Blood and Transplant Analysis 2015[89]
6 to 12 months post-HSCT (cost per cycle)	£3,407.04	
12 to 24 months post-HSCT (cost per cycle)	£1,228.97	
24 months onwards post-HSCT (cost per year)	£525.24	KOL opinion + NHS Schedule of Reference Costs 2018[51]; full blood count + phlebotomy (DAPS05 + DAPS04), biochemistry profile + phlebotomy (DAPS08 + DAPS04), consultant haematologist follow-up outpatient visit (WF01A). Assumes 3 outpatient visits + sets of tests per year.

Based on Table 2 in Appendix I of the CS.[50]
CS = company submission, EFS = event-free survival, HSCT = haemopoietic stem cell transplant, KOL = key opinion leader, NHS = National Health Service

ERG comment: The sources used, and assumptions made, for the estimation of the costs following an HSCT after the procedure itself are appropriate according to the ERG.

Relapsed/progressed disease

Costs for relapsed and progressed disease were split into the following categories: health care resource utilisation (HCRU) costs, monitoring costs, transfusion costs, and treatment costs. All these costs are summarised in Table 5.19 to 5.21. HCRU costs, monitoring costs, and blood transfusion costs were assumed to be the same for AML and MDS patients, based on KOL input. The exception to this was for inpatient stay cost, for which AML and MDS-specific costs were used as sourced from the NHS Schedule of Reference Costs. The frequencies and durations that were used to calculate HCRU costs, Monitoring costs, and Transfusion costs were sourced from TA399[43] (i.e. AML patients only), and costed using the PSSRU 2018. Transfusion costs, for which no cost data were available for the current cost year, were sourced from TA399[43], and inflated to 2017/2018 values using the PSSRU HCHS index.

Table 5.19: HCRU costs associated with relapsed AML/MDS

Resource	Frequency/ cycle	Minutes/ frequency	Staff costs/ minute	Total cost/ cycle	Source (costs)
CNS haematologist	2.62	34.2	£0.90	£80.64	PSSRU 2018: Band 7 Nurse (Nurse, advanced), £54 per working hour
Consultant	1.6	25.6	£1.80	£73.73	PSSRU 2018: Consultant, medical, £105 per working hour
Day care nurse	3.47	40.72	£1.50	£211.95	PSSRU 2018: Band 5 nurse (Staff nurse), £90 per hour of patient contact, £37 per working hour
Day care specialist registrar (SpR)	2.95	22	£0.72	£46.51	PSSRU 2018: Registrar group, £43 per hour
District nurse	0.59	15	£1.50	£13.28	PSSRU 2018: Band 5 nurse (Community-based nurse), £90 per hour of patient contact,

Resource	Frequency/ cycle	Minutes/ frequency	Staff costs/ minute	Total cost/ cycle	Source (costs)
					£37 per working hour
Doctor	0.88	12.67	£1.75	£19.51	PSSRU 2018: Associate specialist, £105 per working hour
Junior Doctor	2.64	9	£0.53	£12.67	PSSRU 2018: Foundation doctor 2, £32 per working hour
Pharmacist	0.42	13.5	£0.73	£4.16	PSSRU 2018: Pharmacist, £44 per working hour
Oncology nurse	0.59	6	£1.85	£6.55	PSSRU 2018: Oncology nurse, £111 per hour of patient contact
Inpatient day (AML)	2.61	1,440	N/A	£413.69	Weighted average of NHS Schedule of Reference Costs[51]: Elective inpatient excess bed days, codes SA25G, SA25H, SA25J, SA25K, SA25L, SA25M
Inpatient day (MDS)	2.61	1,440	N/A	£483.92	Weighted average of NHS Schedule of Reference Costs: Elective inpatient excess bed days, codes SA06G,

Resource	Frequency/ cycle	Minutes/ frequency	Staff costs/ minute	Total cost/ cycle	Source (costs)
					SA06H, SA06J, SA06K

Based on Table 3 in Appendix I of the CS.[50]; source for frequency: TA399[43]
AML = acute myeloid leukaemia; CNS = clinical nurse specialist; CS = company submission; MDS = myelodysplastic syndrome; N/A = not applicable; SpR = specialist registrar; TA = technology appraisal

Table 5.20: Monitoring costs associated with relapsed AML/MDS

Monitoring test	Frequency /cycle	Cost/ frequency	Total cost/ cycle	Source (costs)
Bone marrow aspirates	0.16	£1.11	£1.27	NHS Schedule of Reference Costs: DAPS04
Bone marrow biopsies	0.03	£1.11	£1.14	NHS Schedule of Reference Costs: DAPS04
Peripheral blood smears	0.74	£2.51	£3.25	NHS Schedule of Reference Costs: DAPS05
Blood tests	8.33	£2.51	£10.84	NHS Schedule of Reference Costs: DAPS05
DNA and RNA extractions for molecular testing	0.15	£1.11	£1.26	NHS Schedule of Reference Costs: DAPS04
Extractions for cytogenetic testing	0.13	£12.02	£12.15	NHS Schedule of Reference Costs: DAPS01
Serum blood chemistry	7.74	£1.11	£8.85	NHS Schedule of Reference Costs: DAPS04

Based on Table 4 in Appendix I of the CS[50]; source for frequency and test types: TA399[43]
AML = acute myeloid leukaemia; CS = company submission, DNA = deoxyribonucleic acid, MDS = myelodysplastic syndrome; NHS = National Health Service; RNA = ribonucleic acid; TA = technology appraisal

Table 5.21: Blood transfusion costs associated with relapsed AML/MDS

Blood transfusions	Frequency/ cycle	Cost/ transfusion	Cost/cycle	Source
RBC	4.78	£125.54	£600.08	TA399[43]
Platelets	5.85	£199.00	£1,164.15	

Based on Table 5 in Appendix I of the CS[50]
AML = acute myeloid leukaemia; CS = company submission, MDS = myelodysplastic syndrome; RBC = red blood cells; TA = technology appraisal

In the absence of empirical trial data on the treatment costs for patients with relapsed AML and MDS, treatment regimens for early (<1 year) and late (≥1 year) "relapsers" were determined by KOL input. These are shown below in Table 5.22.

Table 5.22: Relapsed/progressed disease treatment costs

Relapse stage	Treatment regimen	Usage (%)	Treatment used	Source (treatment regimen + uptake)	Source (treatment regimen protocol)
Early (<1 year)	Hypomethylating agents	33.33%	Azacitidine	KOL opinion	South East London Cancer Network: Azacitidine for MDS, CMML and AML
	Salvage chemotherapy	33.33%	MEC	KOL opinion	South East London Cancer Network: MEC (mitoxantrone, etoposide & cytarabine) for relapsed / refractory AML
	Palliative chemotherapy	33.33%	Hydroxycarbamide	KOL opinion	South East London Cancer Network: Hydroxycarbamide (hydroxyurea) for CML and palliative treatment for patients with AML
Late (≥1 year)	Flag/IDA	50.00%	Flag/IDA	KOL opinion	South East London Cancer Network: FLAG+/-Ida (fludarabine, cytarabine, GCSF & idarubicin) for AML
	Second HSCT	50.00%	As initial treatment (treosulfan or busulfan)	KOL opinion	MC-FLUDT14/L Trial III, phase III clinical trial report[90]

Based on Table 6 in Appendix I of the CS[50]
AML = acute myeloid leukaemia; CMML = chronic myelomonocytic leukaemia; CS = company submission; Flag/IDA = fludarabine/cytarabine/granulocyte-colony stimulating factor/idarubicin; HSCT = haematopoietic stem cell transplant; KOL = key opinion leader; MDS = myelodysplastic syndrome; MEC = mitoxantrone, etoposide & cytarabine

The dosages and treatment sequences for each regimen were sourced from protocols used by the South East London Cancer Network. Due to the large number of providers/vial sizes available for some treatments, no mixing and matching of vials to minimise wastage was performed. Instead, the model user can select either the lowest-cost single formulation, the average cost of formulations, or the highest-

cost single formulation. In the base-case analysis, the lowest-cost single formulation of each chemotherapy drug was chosen.

Summaries of the cost per treatment regimen for relapsed patients are shown below in Table 5.23.

Table 5.23: Costs per treatment regimen for relapsed AML/MDS patients

Relapse stage	Regimen	Cost/ treatment cycle	Treat-ment cycles / year	Cost per model cycle	Source (costs)	Source (treatment cycles)
Early (<1 year)	Azacitidine 75 mg/ m^2 /day for 7 days	£1,605.00	6	£740.77	BNF 2019[91]	South East London Cancer Network: Azacitidine for MDS, CMML and AML
	MEC (mitoxantrone, etoposide & cytarabine)	£731.27	2	£112.50	BNF 2019[91]	South East London Cancer Network: MEC for relapsed / refractory AML protocol
	Hydroxycarbamide	£17.59	1	£2.71	BNF 2019[91]	South East London Cancer Network: Hydroxycarbamide (hydroxyurea) for CML and palliative treatment for patients with AML
Late (≥1 year)	Flag/IDA	£3,711.54	2		BNF 2019[91]	South East London Cancer Network: Flag+/-Ida for AML protocol
	Second HSCT (treosulfan or busulfan)	£46,413	One-off	£46,413	MC-FLUDT14/L Trial III, phase III clinical trial report[90]; BNF 2019[91]	Assumption

Based on Table 7 in Appendix I of the CS[50]
AML = acute myeloid leukaemia; BNF = British National Formulary; CMML = chronic myelomonocytic leukaemia; CS = company submission; CSR = clinical study report; Flag/IDA = fludarabinecytarabine/granulocyte-colony stimulating factor/idarubicin; HSCT = haematopoietic stem cell transplant; KOL = key opinion leader; MDS = myelodysplastic syndrome; MEC = mitoxantrone, etoposide & cytarabine

Supportive medications (i.e. antifungals and antimicrobials) were not included, as these were dependent on local protocols. Treatment cycles/year refers to how many treatment cycles patients would undertake within the period they were relapsed, based on KOL input. The costs for HSCT follow-up were not implemented as for the original HSCT to prevent double-counting of monitoring costs and resource use.

Mortality costs

A cancer-related mortality event cost of £12,066 was applied to all patients transitioning to the death state within the model, sourced from 2017/18 PSSRU data.

Summary of health state costs

A list of the health states and associated costs in the economic model is provided below in Table 5.24.

Table 5.24: List of health states and associated costs in the economic model

Health states	Items	Value	Signpost (location in ERG report)
EFS: 6 to 12 months post-HSCT (cost per cycle)	Total cost per cycle	£4,959.47	5.2.9.2.1
EFS: 7 to 12 months post-HSCT (cost per cycle)	Total cost per cycle	£3,407.04	5.2.9.2.1
EFS: 12 to 24 months post-HSCT (cost per cycle)	Total cost per cycle	£1,228.97	5.2.9.2.1
EFS: 24 months onwards post-HSCT (cost per year)	Total cost per cycle	£40.40	5.2.9.2.1
Relapsed / progressed AML: Early relapse (≤12 months)	HCRU costs per cycle	£1,548.72	5.2.9.2.2
	Monitoring costs per cycle	£38.75	5.2.9.2.2
	Blood transfusion costs per cycle	£1,764.24	5.2.9.2.2
	Relapse treatment (pharmacy) costs per cycle	£285.37	5.2.9.2.2
	Total	£3,637.07	5.2.9.2.2
Relapsed / progressed MDS: Early relapse (≤12 months)	HCRU costs per cycle	£1,732.04	5.2.9.2.2
	Monitoring costs per cycle	£38.75	5.2.9.2.2
	Blood transfusion costs per cycle	£1,764.24	5.2.9.2.2
	Relapse treatment (pharmacy) costs per cycle	£285.37	5.2.9.2.2
	Total	£3,820.39	5.2.9.2.2
Relapsed / progressed AML: Late relapse (>12 months)	HCRU costs per cycle	£1,548.72	5.2.9.2.2
	Monitoring costs per cycle	£38.75	5.2.9.2.2
	Blood transfusion costs per cycle	£1,764.24	5.2.9.2.2
	Relapse treatment (pharmacy) costs per cycle	£285.50	5.2.9.2.2
	HSCT cost (one-off)	£46,413.11	5.2.9.2.2
	Total (excluding HSCT)	£3,637.21	5.2.9.2.2
Relapsed / progressed MDS: Late relapse (>12 months)	HCRU costs per cycle	£1,732.04	5.2.9.2.2
	Monitoring costs per cycle	£38.75	5.2.9.2.2

Health states	Items	Value	Signpost (location in ERG report)
	Blood transfusion costs per cycle	£1,764.24	5.2.9.2.2
	Relapse treatment (pharmacy) costs per cycle	£285.50	5.2.9.2.2
	HSCT cost (one-off)	£46,413.11	5.2.9.2.2
	Total (excluding HSCT)	£3,820.53	5.2.9.2.2
Death	Event cost	£12,066.00	5.2.9.2.3

Based on Table 8 in Appendix I of the CS (revised version from response to clarification questions)[50]
AML = acute myeloid leukaemia; CS = company submission; EFS = event free survival; HCRU = healthcare resource utilisation; MDS = myelodysplastic syndrome; PSSRU = Personal Social Services Research Unit

ERG comment: In the absence of trial data, estimates for the utilisation of health care resources, costs of monitoring, and costs of blood transfusion were sourced from NICE TA399.[43] Since TA399 only considered patients with AML, KOL input was sought by the company to confirm that these costs are the same in MDS (except for AML- and MDS-specific unit costs for an inpatient stay). During the clarification phase, the table that provides a 'list of health states and associated costs in the economic model' was revised (i.e. an erroneous entry for a second HSCT in the first year in the relapsed health state was removed) to indicate that a second HSCT may only occur during the second year after the first HSCT procedure.

Despite the information as provided in Table 7 in Appendix I of the CS,[50] which mentions separate costs for a second HSCT for treosulfan versus busulfan patients, for the costs of the second HSCT, an average is used for treosulfan and busulfan patients (i.e. as correctly shown in Table 8 in Appendix I of the CS[50]).

Despite a statement in the model (cell D40, sheet Health_state_costs) 'HSCT costs are only applied once to new late relapsers, i.e. only those patients who have relapsed in the current model cycle' the costs for a second HSCT are implemented by applying them from the start of the second year onwards to all patients in the relapsed health state that die. Since it is not possible in a model such as the current one to calculate the number of patients 'newly' arriving in a health state, the ERG considers the current implementation as a reasonable approximation of what was originally intended.

In their response to the ERG's clarification questions, the company pointed out that the proportions of relapsed patients receiving the various treatments were based on the opinion from the KOL who was unsure, but agreed to the suggestions of the interviewers. Confirmation for this was provided by a second clinical expert. The ERG notes that changing the assumptions regarding the proportions of patients receiving a second HSCT only leads to minor changes in the ICER, and does not change the overall conclusions of the economic analysis.

5.2.9.3 Adverse event costs

For GvHD costs, two French studies (Robin et al. 2017[92] and Esperou et al. 2004[93]) and one US study (Khera et al. 2014[94]) were identified through targeted review. The Esperou study was not considered up to date, therefore only the Robin and Khera studies were included in the model, with the former being included in base-case as French costs were considered more generalisable to a UK population. Costs were converted using a purchasing power parity estimate (PPP) from the Organisation

for Economic Co-operation and Development (OECD), and inflated using the PSSRU HCHS pay and prices index. To prevent overestimation of GvHD costs from the Robin study, excess costs for cGvHD and stage II-IV aGvHD (i.e. the difference between groups with/without cGvHD, and patients with stage 0/I aGvHD and stage II-IV aGvHD) were applied.

For all other adverse events included in the model, NHS reference costs were used as a primary source of duration and cost data, with prior technology appraisals used to help select appropriate HRG codes. Elective inpatient codes were used since AEs were observed within the first 28 days post-HSCT and the average duration of an HSCT procedure inpatient stay (based on the HRG codes for elective inpatients) was around 27 days, thereby assuming that AEs are treated within the same episode of inpatient care.

For costing investigations such as increased GGT and febrile neutropenia HRG codes used in TA523 were not available since the 2014/2015 version of the NHS reference costs. Therefore, the most recent version of the NHS Tariffs including these HRG codes (2016/2017) were considered a more appropriate source of event costs. Elective inpatient costs were unavailable from 2016/2017 tariffs. Therefore, non-elective inpatient costs were used.

For maculopapular rash events, clinical expert opinion confirmed that patients would not require an extended length of stay in hospital, and that patients would most likely receive treatment with systemic steroids. As such, maculopapular rash events were costed assuming a 10 mg daily dose of oral prednisolone and an event duration of 16 days based on the phase III clinical trial. Adverse event costs are summarised in Table 5.25.

Table 5.25: Costs of adverse events

Adverse event	Event cost	
	Mean (SE)	**Source**
Extensive cGvHD	£5,823.68 (N/A)†	Robin et al. 2017[92]
Stage III/IV aGvHD	£14,793.04 (N/A)†	Robin et al. 2017[92]
Mucositis oral	£2,021.36 (£404.27)*	NHS reference costs 2017/18[51] Weighted average of elective inpatient currency codes CB02A-CB02F (Non-Malignant, Ear, Nose, Mouth, Throat or Neck Disorders, with Interventions).
Nausea	£2,263.78 (£452.76)*	NHS reference costs 2017/18[51] Weighted average of elective inpatient currency codes FD10A-FD10M (Non-Malignant Gastrointestinal Tract Disorders).
Diarrhoea	£2,263.78 (£452.76)*	NHS reference costs 2017/18[51] Weighted average of elective inpatient currency codes FD10A-FD10M (Non-Malignant Gastrointestinal Tract Disorders).
Vomiting	£2,263.78 (£452.76)*	NHS reference costs 2017/18[51] Weighted average of elective inpatient currency codes FD10A-FD10M (Non-Malignant Gastrointestinal Tract Disorders).

Adverse event	Event cost	
	Mean (SE)	Source
GGT increased	£983.58 (£196.72)*	NHS National Tariff Payment System 2016/17[95] PA48B Blood Cell Disorders without CC (Admitted Patient Care & Outpatient Procedures): Non-elective spell tariff Inflated to 2017/18 using PSSRU HCHS index[42]
Alanine aminotransferase increased	£983.58 (£196.72)*	NHS National Tariff Payment System 2016/17[95] PA48B Blood Cell Disorders without CC (Admitted Patient Care & Outpatient Procedures): Non-elective spell tariff. In line with Tremblay et al. 2018 and TA523 (HRG code not in 2017/18 tariff). Inflated to 2017/18 using PSSRU HCHS index
Aspartate aminotransferase increased	£983.58 (£196.72)*	NHS National Tariff Payment System 2016/17[95] PA48B Blood Cell Disorders without CC (Admitted Patient Care & Outpatient Procedures): Non-elective spell tariff. Inflated to 2017/18 using PSSRU HCHS index
Blood bilirubin increased	£983.58 (£196.72)*	NHS National Tariff Payment System 2016/17[95] PA48B Blood Cell Disorders without CC (Admitted Patient Care & Outpatient Procedures): Non-elective spell tariff. Inflated to 2017/18 using PSSRU HCHS index
Investigations (other)	£983.58 (£196.72)*	NHS National Tariff Payment System 2016/17[95] PA48B Blood Cell Disorders without CC (Admitted Patient Care & Outpatient Procedures): Non-elective spell tariff. Inflated to 2017/18 using PSSRU HCHS index
Febrile neutropenia	£3,669.67 (£733.93)*	NHS National Tariff Payment System 2016/17[95] PA45Z Febrile Neutropenia with Malignancy (Admitted Patient Care & Outpatient Procedures): Non-elective spell tariff. Inflated to 2017/18 using PSSRU HCHS index
Sepsis	£3,662.50 (£732.50)*	NHS reference costs 2017/18[51] Weighted average of elective inpatient currency codes WJ06A-WJ06J (Sepsis).
Lung infection	£2,924.23 (£584.85)*	NHS reference costs 2017/18[51] Weighted average of elective inpatient currency codes DZ11K-N;P-V (Lobar, Atypical or Viral Pneumonia).
Anorexia	£2,036.45 (£407.29)*	NHS reference costs 2017/18[51] Weighted average of elective inpatient currency codes FD04A-E (Nutritional Disorders).
Syncope	£1,256.05 (£251.21)*	NHS reference costs 2017/18[51] Weighted average of elective inpatient currency codes EB08A-E (Syncope or Collapse).
Rash maculopapular	£21.44 (£4.29)*	BNF. Assumption of 10 mg daily dose of oral prednisolone.

Adverse event	Event cost	
	Mean (SE)	Source
Based on Table 10 in Appendix I of the CS[50] * SE assumed to be 20% of the mean † Not applicable as these estimates were derived as the difference in costs for patients with/without cGvHD and with stage 0/I and stage II-IV aGvHD respectively aGvHD = acute graft vs host disease, BNF = British National Formulary, cGvHD = chronic graft vs host disease, CS = company submission, GGT = gamma glutamyl transferase, HCHS = Hospital and Community Health Services, N/A = not applicable, NHS = National Health Service, SE = standard error		

ERG comment: The company has costed AEs using NHS reference costs if possible, and using NHS tariffs otherwise. The ERG considers this as an appropriate costing strategy for AEs. The impact of the costs of AEs on the cost effectiveness results are small.

6. Cost effectiveness results

6.1 Company's cost effectiveness results

The discounted base-case results indicated that treosulfan generated more QALYs than busulfan at lower costs. Therefore, treosulfan was dominant over busulfan. The full (discounted) base-case results are presented in Table 6.1.

Table 6.1: Company base-case cost effectiveness results (discounted)

Technologies	Total costs (£)	Total LYGs	Total QALYs	Incr. costs (£)	Incr. LYGs	Incr. QALYs	ICER versus baseline (£/QALY)
Treosulfan	£137,062	8.73	6.44				
Busulfan	£160,821	7.71	5.55	£23,759	-1.01	-0.89	Dominated

Based on Table 55 of the CS[1]
CS = company submission; ICER = incremental cost effectiveness ratio; Incr. = incremental; LYGs = life years gained; QALYs = quality-adjusted life years

The disaggregated discounted results by health state are given in Table 6.2. It can be seen that, compared to busulfan, treosulfan approximately halved the time in the relapse/progression health state. Since the costs of the relapse/progression health state are high, the relapse/progression costs saved by treosulfan outweighed the lower busulfan treatment and EFS costs. Hence, the dominance.

Table 6.2: Disaggregated (discounted) results by health state

Health state	Costs		QALYs		LYGs	
	Treosulfan	Busulfan	Treosulfan	Busulfan	Treosulfan	Busulfan
Event-free survival	£102,990	£100,799	6.137	4.967	8.168	6.629
Relapse / progression: AML	£15,898	£31,800	0.193	0.372	0.357	0.691
Relapse / progression: MDS	£9,468	£18,910	0.109	0.210	0.202	0.391
Dead	£8,705	£9,312	-0.004	-0.007	NA	NA
Total	£137,062	£160,821	6.44	5.55	8.726	7.711

Based on Table 56 of the CS[1] and the electronic model
AML = acute myeloid leukaemia; CS = company submission; LYG = life years gained; MDS = myelodysplastic syndrome; QALYs = quality-adjusted life-years.

6.2 Company's sensitivity analyses

6.2.1 Probabilistic sensitivity analysis

The probabilistic sensitivity analysis (PSA) conducted by the company was based on 5,000 iterations. The input parameters included in the PSA were reported in Appendix L of the company submission.[18] Survival parameters, including all Cholesky matrices for all survival functions and patients subgroups, were provided in Appendix M2,[96] M3[97] and M4[98] of the CS. Wherever possible, standard errors for the

input parameters were obtained from the corresponding source or derived from other measures (e.g. standard deviations, 95% confidence intervals, etc.). If no uncertainty range was available, standard errors were assumed as 20% of the mean value.

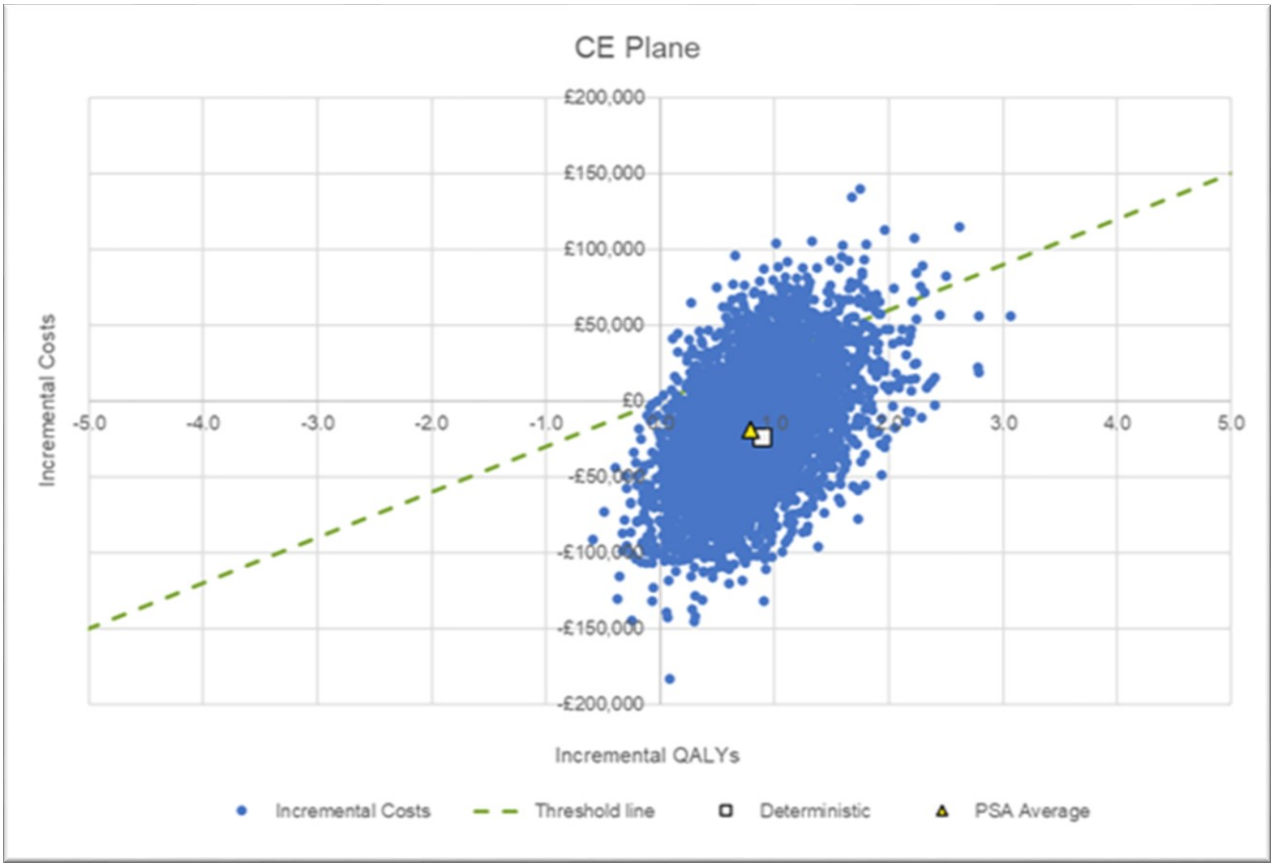
The discounted results from the PSA are shown in Table 6.3. The incremental costs and incremental QALYs obtained from the PSA were plotted in the cost effectiveness (CE) plane, from which a cost effectiveness acceptability curve (CEAC) was derived. These are shown in Figures 6.1 and 6.2, respectively. The probabilistic results were in line with the deterministic ones since treosulfan was also dominant over busulfan. The vast majority of the 5,000 iterations provided results in the eastern quadrants of the CE plane, where treosulfan is more effective than busulfan. In particular, 69.6% of the iterations were in the south-eastern quadrant of the CE plane, where treosulfan dominates busulfan. The CEAC also showed that treosulfan has an 84.3% probability of being cost effective at a threshold of £20,000 per QALY and an 89.9% probability of being cost effective at a threshold of £30,000 per QALY.

Table 6.3: Company base-case probabilistic cost effectiveness results (discounted)

Technologies	Total costs (£)	Total QALYs	Incremental costs (£)	Incremental QALYs	ICER (£)
Treosulfan	NR	NR			
Busulfan	NR	NR	£19,084	-0.79	Dominated

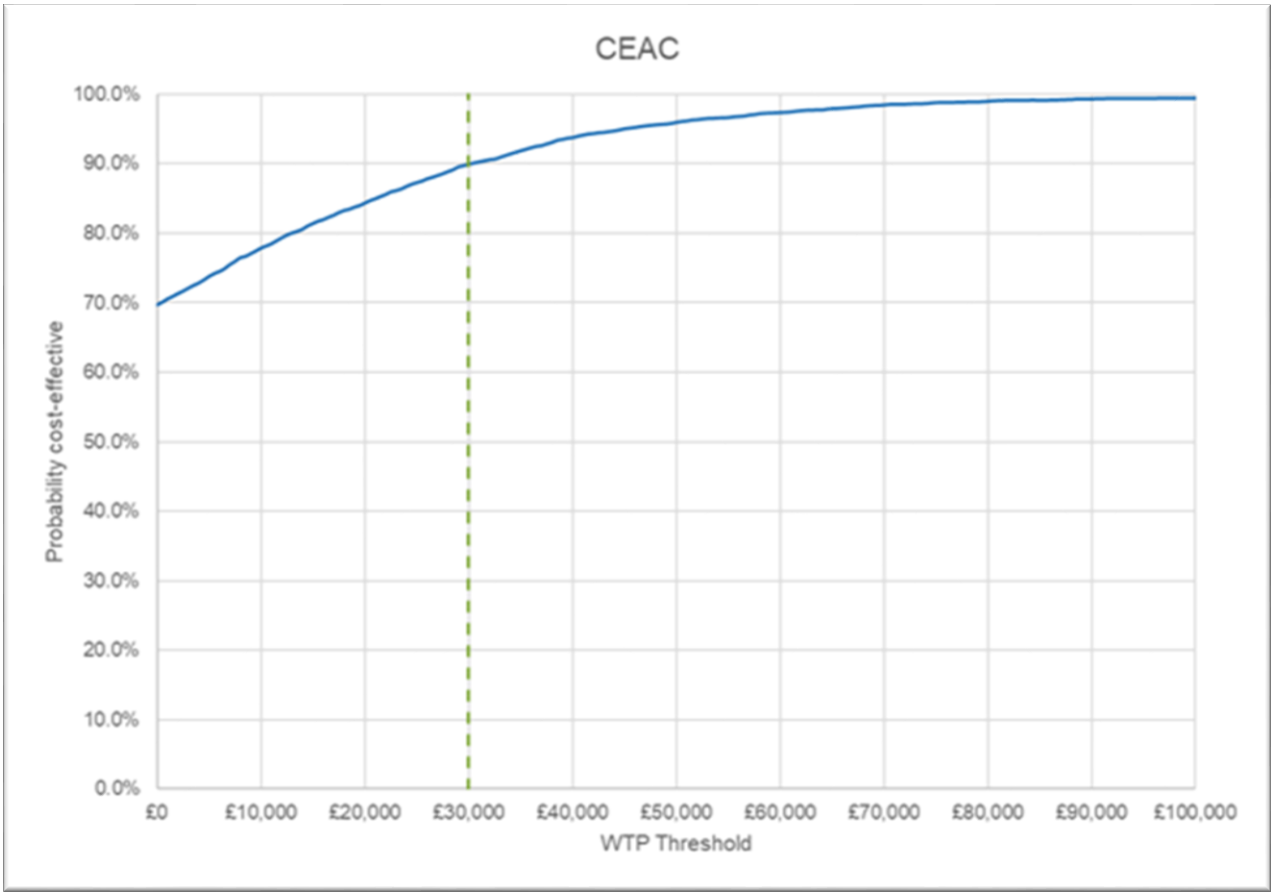
Based on page 176 of the CS[1]
CS = company submission; ICER = incremental cost effectiveness ratio; QALY = quality-adjusted life year; NR = not reported

Figure 6.1: Scatterplot form the probabilistic sensitivity analysis iterations



Based on Figure 24 of the CS[1]
CS = company submission; PSA = probabilistic sensitivity analysis; QALY = quality-adjusted life year

Figure 6.2: Cost effectiveness acceptability curve

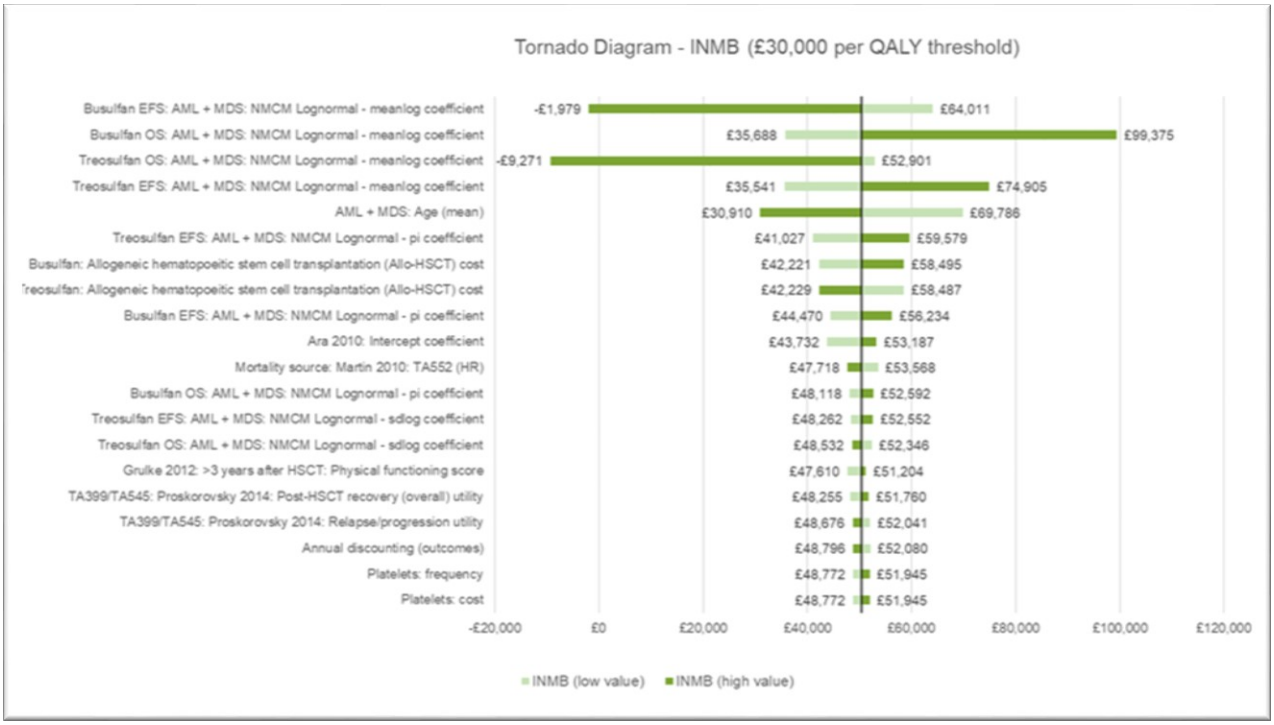


Based on Figure 25 of the CS[1]
CEAC = cost effectiveness acceptability curve; CS = company submission; WTP = willingness to pay

6.2.2 Deterministic sensitivity analysis

The company also conducted a one-way deterministic sensitivity analysis. The value of each parameter included in the analysis was varied by +/- 20% with respect to the base-case value, while the remaining parameters were kept constant at their base-case values. The tornado plot in Figure 6.3 shows the impact on the incremental net monetary benefit (INMB), calculated using a £30,000 per QALY willingness-to-pay (WTP) threshold, of the 20 parameters which caused the largest changes in the INMB. From this diagram, it is clear that only changes in the meanlog coefficients of the busulfan EFS distribution and the treosulfan OS distribution resulted in a negative INMB or, equivalently, in an ICER above £30,000 per QALY. The company also presented the results of the one-way deterministic sensitivity analysis in terms of the ICER (not shown here). This confirmed the conclusions based on the INMB but also showed that for all parameters the ICERs were negative, indicating that treosulfan was dominating busulfan. The only exceptions (with positive ICERs) were the meanlog coefficients of the treosulfan OS distribution, the busulfan EFS distribution and the busulfan OS distribution, which resulted in upper bound ICERs of £34,821, £33,040 and £20 per QALY, respectively.

Figure 6.3: Tornado diagram – company's preferred assumptions



Based on Figure 27 of the CS[1]
AML = acute myeloid leukaemia; CS = company submission; EFS = event-free survival; INMB = incremental net monetary benefit; NMCM = non-mixture-cure models; MDS = myelodysplastic syndrome; QALY = quality-adjusted life year

ERG comment: The tornado diagram indicated that only the coefficients of the survival distributions might have some impact on decision uncertainty. While this might be the case, it should be noted that the one-way sensitivity analysis was conducted, allowing 20% variation from mean value for all parameters, which seems arbitrary and may not represent an equally plausible range of variation for all input parameters. Similar to the PSA described in the CS, the limits of (95%) confidence intervals should be used for each parameter to calculate the upper and lower bounds of the bars in the tornado diagram. In their response to clarification question C29, the company clarified that the choice of 20% variation for all parameters for the deterministic sensitivity analysis was chosen as the objective of the deterministic sensitivity analysis was to identify the most sensitive parameters in the model, and that they were concerned about producing misleading results from the DSA using 95% confidence intervals due to the covariance between survival function parameters. Furthermore, they cited that the choice of 20% variation for inputs for the deterministic sensitivity analysis was consistent with two recent AML appraisals (TA399, TA523).

While the ERG acknowledge the concern that covariance between survival function parameters may affect 95% confidence intervals, this argument only applies to a small number of parameters included in the DSA and the ERG still feel that the confidence interval approach represents better practice. This is because, while the +/- 20% approach undertaken by the company provides an idea of which parameters are having the biggest impact on model results, the results of the company tornado plot do not provide information on the impact of the actual uncertainty related to each parameter, which would be provided if an approach based on confidence intervals were used. Unfortunately, given the time

constraints associated with this project, the ERG was not able to test the impact of using 95% confidence intervals for the deterministic sensitivity analysis. Therefore, while the ERG agrees that the model results seem to be more sensitive to changes in the survival distributions, the ERG also considers that the tornado diagram presented above should be interpreted with caution.

6.2.3 Scenario analyses

The company undertook a number of scenario analyses in order to examine the impact of certain assumptions on the model outcomes. The results of the scenarios tested are summarised in Table 6.4. The scenarios with the largest impact on the results were those where a shorter time horizon was assumed and alternative survival curves were considered. Nevertheless, in all scenarios, treosulfan was dominant over busulfan, indicating that the model is robust to changes in these assumptions.

ERG comment: All the scenario analyses conducted by the company resulted in treosulfan dominating busulfan. Even in the scenario that was the least favourable for treosulfan, when a five-year time horizon was assumed, treosulfan still generated more QALYs than busulfan (0.29) at lower costs (-£6,285). Based on the base-case, sensitivity and scenario analyses presented by the company, the ERG considers that the results of the model are very robust and that any additional scenarios are likely to result in treosulfan dominance. Since the costs associated to the relapse health state are high, even a minor reduction in the time spent in the relapse health state will be sufficient for treosulfan dominance. Despite this, and as mentioned throughout Section 5.2.6 of this report, the ERG believes that potentially important scenarios to test the main structural uncertainties in the economic model were missing in the CS. Therefore, several additional scenarios were conducted by the ERG in Section 7 of this report.

Table 6.4: Scenario analyses conducted by the company

Scenario	Alternative input	Base-case value	Parameter value in scenario	Incremental costs (£)	Incremental QALYs	ICER (£)
Base-case				-£23,759	0.89	Treosulfan dominates
1	Utilities from Grulke et al. 2012[71] and TA399[43]. Mapping from Proskorovsky et al. 2014[75]	Utilities from Grulke et al. 2012[71] and TA399[43]. Mapping from McKenzie et al. 2009[76]	See Table 5.16 – section "Utility scenario 1" for unadjusted (for age and sex) utility values	-£23,759	0.79	Treosulfan dominates
2	Utilities from Castejon et al. 2018[59] with serious AE disutilities from Stein et al. 2018[62]	Utilities from Grulke et al. 2012[71] and TA399[43]	See Table 5.16 – section "Utility scenario 2" for unadjusted (for age and sex) utility values	-£23,759	0.96	Treosulfan dominates
3	MDS specific utility for RBC transfusion dependent patients from Szende et al. 2009[64]	Assumed equal to AML relapse value from TA399[43]/ TA545[44] using Proskorovsky et al. 2014[75] mapping algorithm	0.535	-£23,759	0.89	Treosulfan dominates
4	Relapse disutility multiplier from Proskorovsky et al. 2014[75] mapping of TA399[43] utilities	Relapse disutility multiplier from Proskorovsky et al. 2014[75] mapping of TA399[43] utilities applied to post-	0.623 multiplier applied to 0.660 discharge utility	-£23,759	0.92	Treosulfan dominates

Scenario	Alternative input	Base-case value	Parameter value in scenario	Incremental costs (£)	Incremental QALYs	ICER (£)
	applied to discharged patients	HCST recovery ≤6 months				
5	Busulfan dose	4 x 0.8 mg/kg/day (as per the trial population)	Single 3.2 mg/kg/day	-£22,614	0.89	Treosulfan dominates
6	Separate modelling of AML and MDS	AML + MDS patients pooled to generate combined OS and EFS	OS and EFS modelled separately for AML and MDS patients. Overall results as weighted average (63.88% AML and 36.12% MDS)	-£25,363	0.87	Treosulfan dominates
7	Shorter time horizon	40 years	5 years	-£6,285	0.29	Treosulfan dominates
8	Shorter time horizon	40 years	10 years	-£13,761	0.55	Treosulfan dominates
9	Alternative survival curves	NMCM lognormal	Gamma	-£15,561	1.09	Treosulfan dominates

Based on Tables 58, 60, 61, 62, 64, 65, 66 and 67 of the CS[99]
AE = adverse event; AML = acute myeloid leukaemia; CS = company submission; EFS = event-free survival; ICER = incremental cost effectiveness ratio, MDS = myelodysplastic syndrome; OS = overall survival; QALY = quality-adjusted life year; RBC = red blood cells; TA = technology appraisal

6.2.4 Subgroup analysis

The company performed subgroup analysis for AML and MDS patients separately. As in the base-case, NMCM lognormal distributions were assumed for OS and EFS for both treatment arms and sub-populations. Cost effectiveness results for AML and MDS patients are summarised in Tables 6.5 and 6.6, respectively. In both subgroups treosulfan was dominant. In the AML subgroup, cost savings for treosulfan were larger than in the base-case, but with smaller QALY gains. The opposite was observed for the MDS subgroup.

Table 6.5: AML subgroup cost effectiveness results

Technologies	Total costs (£)	Total LYGs	Total QALYs	Incr. costs (£)	Incr. LYGs	Incr. QALYs	ICER versus baseline (£/QALY)
Treosulfan	£142,006	8.90	6.54				
Busulfan	£178,633	8.13	5.77	£36,627	-0.76	-0.77	Dominated

Based on Table 68 of the CS[1]
AML = acute myeloid leukaemia; CS = company submission; ICER = incremental cost effectiveness ratio; Incr. = incremental; LYGs = life years gained; QALYs = quality-adjusted life years.

Table 6.6: MDS subgroup cost effectiveness results

Technologies	Total costs (£)	Total LYGs	Total QALYs	Incr. costs (£)	Incr. LYGs	Incr. QALYs	ICER versus baseline (£/QALY)
Treosulfan	£128,063	8.35	6.20				
Busulfan	£133,501	7.03	5.14	£5,172	-1.32	-1.05	Dominated

Based on Table 69 of the CS[1]
CS = company submission; ICER = incremental cost effectiveness ratio; Incr. = incremental; LYGs = life years gained; MDS = myelodysplastic syndrome; QALYs = quality-adjusted life years

ERG comment: The ERG asked the company to explain the rationale for including cost effectiveness subgroup analyses (AML/MDS) in the submission even though these were not mentioned in the scope and to clarify whether the subgroups considered were appropriate or not, since in the CS it was mentioned that the underlying disease is not the "*primary determinant of conditioning regimen*"; risk factors seem to be more important (see clarification question C11).[12] In their response, the company indicated that subgroup analyses were performed for validation purposes, i.e. to ensure that the results were consistent within each population and with the overall results.

The ERG would like to point out that the selection of survival models should be based on subgroup-specific data. Therefore, all the analyses presented in Section 5.2.6 of this report should be repeated for each subgroup. Assuming an NMCM lognormal distribution for OS and EFS for both treatment arms and sub-populations does not seem consistent with the methodology applied by the company (i.e. low AIC/BIC and good visual fit) for the base-case. This is particularly clear for the MDS subgroup where the NMCM lognormal distribution overestimates both OS and EFS for treosulfan (especially the latter) and underestimates (although to a lower extent) both OS and EFS for busulfan. Repeating the steps in Section 5.2.6 would likely result in a different selection of parametric curves.

Furthermore, in the clarification question C25,[12] the ERG asked the company to confirm whether the results of the subgroups analyses were obtained using subgroup-specific patient characteristics and not the pooled characteristics as in the base-case. However, the company indicated that the subgroup analyses used the pooled patient characteristics to minimise heterogeneity. The ERG does not agree with this assumption. Subgroup analyses require subgroup-specific input parameters, including patient characteristics.

6.3 Model validation and face validity check

Validation was performed by two external clinical experts, with experience in the treatment of AML and MDS with HSCT, and one internal expert from the company. These efforts focused on validating long-term survival assumptions (particularly those related with SMRs of HSCT patients), health state utility, cost inputs and post-relapse treatment regimens for AML and MDS. The questions submitted to the experts were provided in Appendix L of the CS.[18]

Furthermore, the overall quality of the economic evaluation was assessed with the Philips' checklist,[100] which was filled in by an independent health economist and provided in Appendix M1 of the CS.[101]

ERG comment: As requested by the ERG in the request for clarification (question C2), the company provided additional details of the communication with clinical experts.[12]

Additionally, since other important aspects of validation were not reported in the CS (e.g. quality control/verification of the calculations in the model), the ERG asked the company to provide details about what validation efforts were performed in Section B.3.8 of the company submission and the results of these validation efforts (clarification question C20).[11]

The company indicated that, in addition to the validation work conducted with the clinical experts, general logic and calculation checks were performed by the model developers and another health economist who were not involved in the model development. The tests performed included logical checks of calculations, stress testing of parameters, and individual setting of inputs to zero to determine whether the expected outputs were generated. The results of these tests were not reported. Therefore, the ERG cannot assess the degree of internal validation of the model. The additional validation efforts conducted by the ERG led to the identification of several modelling errors that are described in Section 7.1.2 of this report.

Finally, in clarification question C9 g, the ERG asked the company to provide probability estimates which can be used to validate the tails of the parametric curves presented in the model.[12] In their response, the company indicated that the tails of the curves were only validated up to five years, which corresponds to the base-case "cure point". The clinical experts consulted by the company considered that using the complete parametric curves to model survival was inappropriate. As mentioned in the critique to Section 5.2.6.1, the clinical experts agreed that the survival curves used in the base-case were plausible.

7. Evidence review group's additional analyses

7.1 Exploratory and sensitivity analyses undertaken by the ERG

7.1.1 Explanation of the company adjustments after the request for clarification

Following the clarification questions from the ERG, the company made the following four amendments to the originally submitted cost effectiveness model:

- Clarification question C21: correcting negative costs appearing at the end of the model time horizon.
- Clarification question C24: age decrement in utility associated to adverse events implemented in the model.

After the changes made by the company, treosulfan still generated more QALYs than busulfan (0.89) at lower costs (-£23,668, compared to -£23,759 in the base-case). Therefore, the effect of these changes on the base-case results was minor.

7.1.2 Explanation of the ERG adjustments

The changes made by the ERG (to the model received with the response to the request for clarification) were subdivided into the following three categories (according to Kaltenthaler et al. 2016)[102]:

- Fixing errors (correcting the model where the company's electronic model was unequivocally wrong).
- Fixing violations (correcting the model where the ERG considered that the NICE reference case, scope or best practice has not been adhered to).
- Matters of judgement (amending the model where the ERG considered that reasonable alternative assumptions are preferred).

After these changes were implemented, additional scenario analyses were explored by the ERG.

7.1.2.1 Fixing errors

1. Overall and event-free survival probabilities were incorrectly calculated using mortality rates instead of transition probabilities. This happened in columns U and V of the model engine sheets. Note that in columns Q and S, the survival probabilities are correct since rates were transformed into transition probabilities. A more detailed explanation is included in Appendix 2 and in the in the ERG revised version of the electronic model.
2. Yearly values where re-scaled to daily values using a factor 1/364. This was used in life tables, the calculation of health state costs and health state utilities. The ERG understands that in the economic model with a cycle of 28 days, a year is assumed to have 364 days. However, for the yearly values obtained from life tables, the calculation of health state costs or health state utilities this assumption does not necessarily hold. Therefore, the ERG prefers a factor 1/365.25 (and then multiply this by 28 to adjust for the cycle length). In any case, this is a minor error with almost no impact on the model results.

7.1.2.2 Fixing violations

3. Subgroup analyses (does not affect the base-case):
    a. Patient characteristics should be subgroup-specific.

b. Goodness of fit of survival curves should be reassessed based on subgroup-specific data.

7.1.2.3 Matters of judgement

4. The company applied mortality rates from HSCT-adjusted life tables to amend EFS, when EFS rates were the lowest. Since EFS is defined by three events (relapse, graft failure or death), the ERG prefers using unadjusted EFS allowing thus OS and EFS curves crossing. A more detailed explanation is included in Appendix 2 and in the in the ERG revised version of the electronic model.
5. Treatment effectiveness:
    a. Modelling OS according to a NMCM Weibull distribution.
6. Life tables:
    a. The most recent UK life tables (2015/2017) were not used.
    b. The calculation of q_x (i.e. the mortality rate between age x and x +1) in the model did not correspond with the formula used in the in the UK life tables.

The main assumptions made by the company and the ERG for their preferred base-case-analyses are summarised in Table 7.1.

Table 7.1: Company and ERG base-case preferred assumptions

Base-case preferred assumptions	Company	Justification	ERG	Justification for change
OS and EFS probability calculations	Calculation based on rates	N/A	Calculation based on transition probabilities	Programming error
Yearly values re-scaled to daily values	1/364	The model assumes 1 year = 364 days	1/365.25	Section 7.1.2.1
OS modelling	NMCM lognormal	Lowest AIC/BIC	NMCM Weibull	Section 5.2.6.1
EFS modelling	NMCM lognormal	Lowest AIC/BIC	NMCM lognormal	Not changed
"Cure point"	5 years	Expert opinion	5 years	Not changed
SMR for HSCT mortality	2.30	Expert opinion and TA552[41]	2.30	Not changed
Long-term mortality	Approach 5 in Section 5.2.6.2	Expert opinion	Approach 5 in Section 5.2.6.2	Not changed
HRQoL	Approach described in Section 5.2.8.4	Described in Section 5.2.8	Approach described in Section 5.2.8.4	Not changed
Resource use and costs	Approach described in Section 5.2.9	Described in Section 5.2.9	Approach described in Section 5.2.9	Not changed
UK life tables	2014/2016	N/A	2015/2017	Most recent version
Calculation of q_x in life tables	Approximation	N/A	Formula in UK life tables	Same formula as in UK life tables

AIC = Akaike information criterion; BIC = Bayesian information criterion; EFS = event-free survival; ERG = Evidence Review Group; HRQoL = health-related quality of life; HSCT = haematopoietic stem cell transplantation ; N/A = not applicable; NMCM =non-mixture cure models; OS = overall survival; q_x = mortality rate between age x and x +1; SMR = standardised mortality ratio

7.1.3 Additional scenarios conducted by the ERG

The ERG conducted additional scenario analyses in which several sources of uncertainty identified by the ERG were explored. These were the uncertainties associated with modelling EFS, OS, the duration of the "cure point", the approach used to model long-term mortality, the SMR used to adjust HSCT background mortality, utilities, costs and resource use. A list of the scenario analyses conducted by the ERG is provided below.

7.1.3.1 Scenario set 1: changing EFS parametric distributions

As explained in Section 5.2.6.1 of this report, the ERG chose the NMCM lognormal distribution to model EFS for its preferred base-case. The ERG explored an additional scenario assuming an MCM lognormal distribution for EFS. Furthermore, since these two distributions seem to overestimate treosulfan EFS, the ERG conducted another scenario assuming a gamma distribution for treosulfan EFS and an NMCM lognormal distribution (as in the base-case) for busulfan EFS, even though choosing different parametric distributions to model different treatment arms is contrary to the advice in DSU 14.[46] This scenario can be useful to estimate the size of the bias in the base-case but its results should be interpreted with caution.

7.1.3.2 Scenario set 2: changing OS parametric distributions

As explained in section 5.2.6.1 of this report, the ERG chose the NMCM Weibull distribution to model OS for its preferred base-case. The ERG explored three additional scenarios assuming MCM Weibull, NMCM lognormal and MCM lognormal distributions for OS.

7.1.3.3 Scenario set 3: changing duration of cure point

As explained in Section 5.2.6.2 of this report, the company included a fixed "cure point" of five years in the model. The rationale for including this "cure point" was that HSCT is a potentially curative treatment. Thus, prior to this "cure point", transitions between the health states of the model are calculated based on the parametric OS and EFS curves. After the "cure point", only background mortality is determining the transitions between the health states. This choice of five years was made based on the opinion of two clinical experts. The impact of choosing different "cure points" on the economic results was explored by the ERG in this series of scenarios.

7.1.3.4 Scenario set 4: alternative approach to modelling long-term mortality (explicit vs. no explicit cure point)

As explained in Section 5.2.6.2 of this report, the company presented in the CS five different approaches to model long-term mortality. In the response to the clarification question C8,[12] the company confirmed that only approaches 3 (use of parametric curves or HSCT-specific life tables – as determined by SMRs – depending on which has the highest mortality rate) and 5 (use of parametric curves up to the fixed "cure point", followed by switch to HSCT-specific life table – as determined by SMRs – mortality rates) are plausible. Both the company and the ERG chose approach 5 for their preferred base-case. The impact of selecting approach 3 to long-term mortality was explored by the ERG in an additional scenario.

7.1.3.5 Scenario set 5: alternative SMRs to model HSCT mortality

As explained in Section 5.2.6.2 of this report, the company presented six different SMRs to calculate long-term mortality post-HSCT. Based on feedback from two clinical experts, the company (and the

ERG) selected an HR = 2.3 from TA552[41] for the base-case. The impact of choosing the other different SMRs on the economic results was explored by the ERG in series of scenarios.

7.1.3.6 Scenario set 6: utilities

In this set of scenarios, the ERG explored the impact of using different sources for the utility estimates on the model results. The first scenario tested the impact of the choice of mapping algorithm on the base-case health state utility values obtained from Grulke et al. 2012[71]. The Grulke values were still used for the health state utility values, but the Proskorovsky et al. 2014[75] mapping algorithm was replaced by the McKenzie et al. 2009[76] algorithm, as this was the algorithm that was considered the second best in this population. The second scenario utilised the DCE values available for health state utility values from Castejon et al. 2018[59] and for adverse events from Stein et al. 2018[62], in order to test the impact of values obtained from DCE on model results. In this scenario, the relapse/progression utilities were applied using the standard age-adjusted utility function, rather than as a disutility multiplier applied to short-term post-HSCT recovery patients (≤6 months), as was done by the company in their DCE utility scenario analysis. The third scenario tests the impact of allowing the relapse utility to vary between AML and MDS patients. This scenario uses the base-case utility values, except for utilising the MDS specific utility value from Szende et al. 2009.[64] The final scenario uses the utility values from TA399[43], using the Proskorovsky et al. 2014[75] mapping algorithm, as these values are established in the previous NICE appraisal and are already used in the base-case for relapse utilities and serious AE disutilities.

7.1.3.7 Scenario set 7: costs and resource use

The ERG has conducted a series of analysis to assess the sensitivity of the results to different assumptions regarding costs. In particular, the proportion of patients that receives, the relatively costly second HSCT procedure in the second year of relapse is varied. In base-case, this value is 50%. In Cost scenario 1 this value is changed to 100%, and in Cost scenario 2 this is changed into 0%. A third cost scenario analysis (Cost scenario 3) was conducted on the influence of the 100% wastage assumption for the administration of treosulfan. In scenario 6.3, this value was changed to 0% (similar to scenario 5 in the submission by the company).

7.1.3.8 Scenario set 8: subgroup analyses

In Section 6.2.4, the company presented results analysis for AML and MDS patients separately. In these analyses, the company assumed the same EFS and OS distributions as in the base-case with the pooled population. As explained in the critique to that section, the ERG considers that the selection of survival models should be based on subgroup-specific data. Therefore, all the analyses presented in Section 5.2.6 of this report were repeated for each subgroup and the results were presented in Appendix 1. Based on these analyses, the ERG concluded that for the AML subpopulation the preferred OS distribution was the MCM lognormal and the preferred EFS distribution was the NMCM lognormal. An additional scenario using a Gamma distribution for EFS was also explored. For the MDS subpopulation the preferred OS distribution was the NMCM Weibull and the preferred EFS distribution was the MCM lognormal. Since this choice for EFS seems to overestimate survival gains for treosulfan, and additional scenario assuming different distributions for each treatment arm (a lognormal for treosulfan and an MCM Weibull for busulfan) was also conducted by the ERG. Note that choosing different parametric distributions to model different treatment arms is contrary to the advice in DSU 14.[46] Nevertheless, this scenario can be useful to estimate the size of the bias in the base-case MDS scenario but its results

should be interpreted with caution. Furthermore, all subgroup analyses were run selecting subgroup-specific patient characteristics (instead of using the characteristics of the pooled population).

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

7.2.1 Results of the ERG preferred base-case scenario

The results of the ERG preferred base-case analysis (as outlined in Section 7.1.2 of this report) are shown in Table 7.2. The implementation of the ERG preferred assumptions resulted in treosulfan generating 0.78 more QALYs than busulfan at lower costs (-£17,689). Therefore, treosulfan dominated busulfan as in the company base-case. In the company base-case, treosulfan generated 0.89 more QALYs than busulfan and saved -£23,759. Thus, in the ERG base-case, both cost savings and QALY gains for treosulfan were smaller than in the company base-case.

Table 7.2: ERG preferred deterministic base-case results (discounted)

Technologies	Total costs (£)	Total LYGs	Total QALYs	Incr. costs (£)	Incr. LYGs	Incr. QALYs	ICER versus baseline (£/QALY)
Treosulfan	£128,147	8.75	6.49	-£17,689	0.92	0.78	Treosulfan dominates
Busulfan	£145,836	7.84	5.71				

ERG = Evidence Review Group; ICER = incremental cost effectiveness ratio, Incr. = incremental, LYGs = life years gained, QALY = quality-adjusted life year

The ERG also conducted a PSA using the ERG preferred base-case assumptions. The results of the ERG PSA are shown in Table 7.3. The probabilistic results are in line with the deterministic ones.

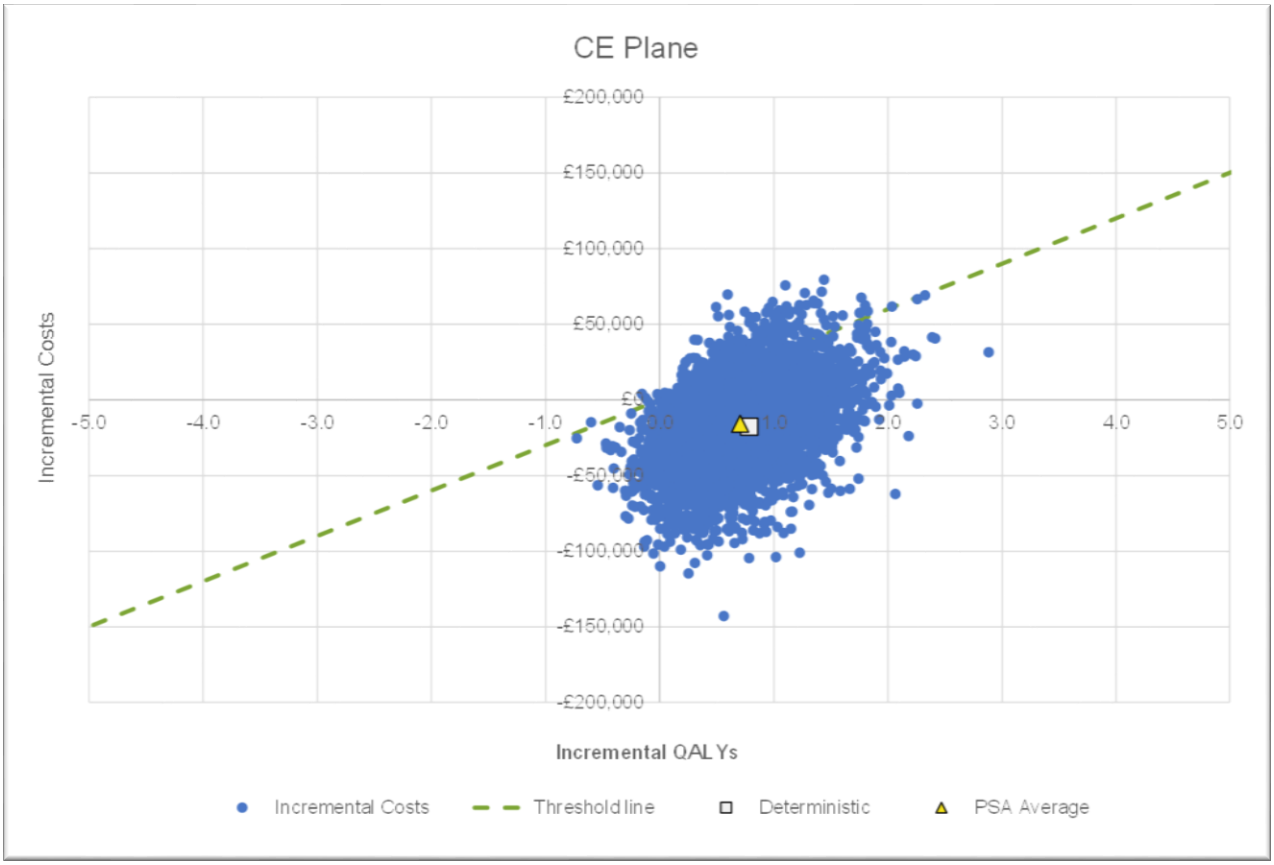
Table 7.3: ERG base-case probabilistic cost effectiveness results (discounted)

Technologies	Total costs (£)	Total QALYs	Incr. costs (£)	Incr. QALYs	ICER (£/QALY)
Treosulfan	NR	NR	-£15,857	0.70	Treosulfan dominates
Busulfan	NR	NR			

ERG = Evidence Review Group; ICER = incremental cost effectiveness ratio, Incr. = incremental, QALY = quality-adjusted life year

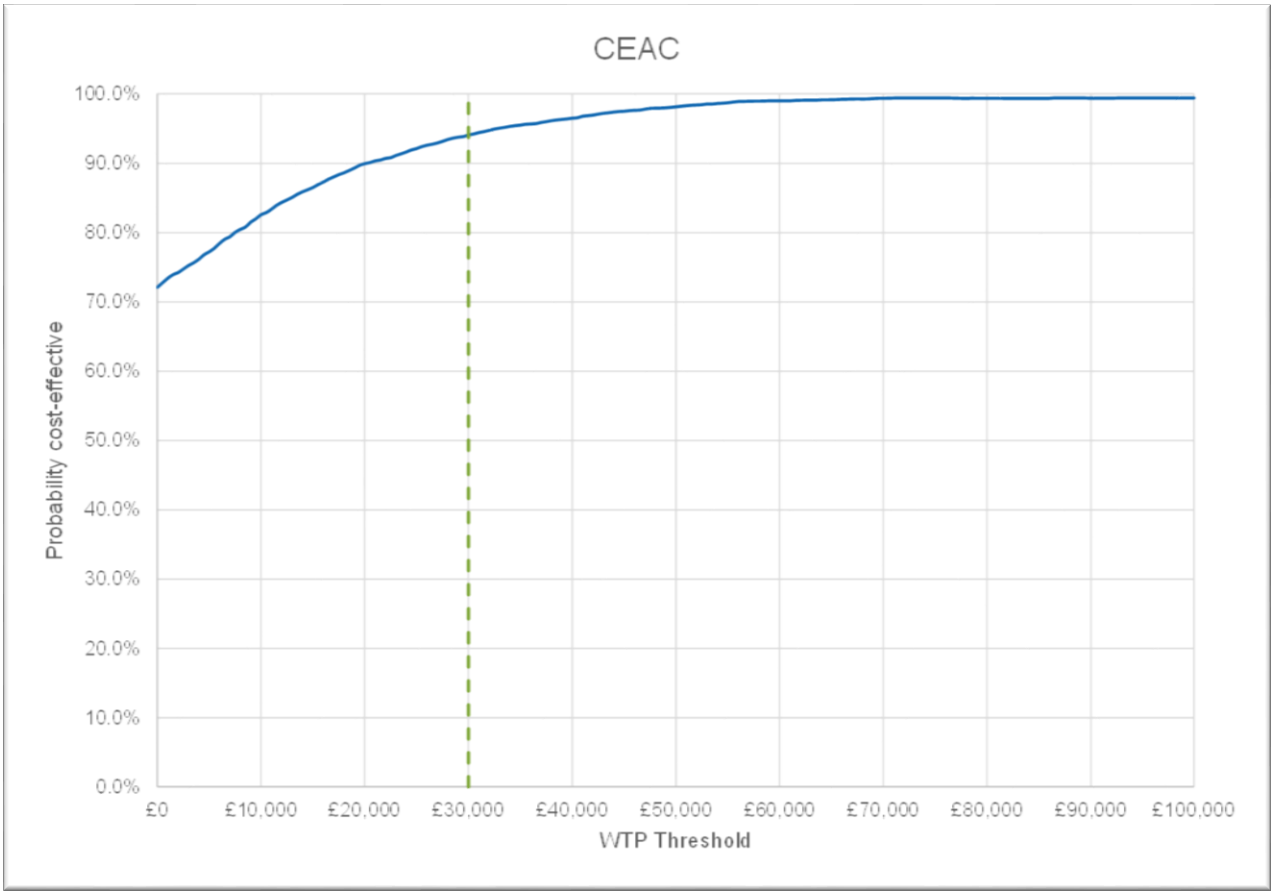
The incremental costs and incremental QALYs obtained from the ERG PSA were plotted in the CE-plane and a CEAC was calculated. These are shown in Figures 7.1 and 7.2, respectively. As in the company base-case, the vast majority of the 5,000 iterations were in the eastern quadrants of the CE plane, where treosulfan is more effective than busulfan. In particular, 72.1% of the iterations were in the south-eastern quadrant of the CE plane, where treosulfan dominates busulfan. The CEAC also showed that treosulfan has an 89.9% probability of being cost effective at a threshold of £20,000 per QALY and a 94.1% probability of being cost effective at a threshold of £30,000 per QALY. These results are in line with those presented in the company base-case.

Figure 7.1: ERG preferred cost effectiveness plane



Based on electronic model
CE = cost effectiveness; ERG = Evidence Review Group; PSA = probabilistic sensitivity analysis; QALY = quality-adjusted life year

Figure 7.2: ERG preferred cost effectiveness acceptability curve



Based on electronic model
CEAC = cost effectiveness acceptability curve; ERG = Evidence Review Group; WTP = willingness-to-pay

7.2.2 Results of the ERG additional exploratory scenario analyses

7.2.2.1 Additional scenario 1: changing EFS parametric distributions

In this series of scenarios, the ERG assessed the impact of using different parametric survival curves for extrapolating of EFS beyond the duration of MC-FludT.14/L Trial II. In particular, the ERG assumed an MCM lognormal distribution for EFS and in a second scenario a gamma distribution for treosulfan EFS and an NMCM lognormal distribution for busulfan EFS were assumed. The purpose of the second scenario was to estimate the size of the bias in the results when using NMCM/MCM lognormal distributions since these seem to overestimate EFS for treosulfan. However, choosing different parametric distributions to model different treatment arms is not recommended.[46] The results in Table 7.4 indicate that the model is robust to the changes made (treosulfan still dominates busulfan in both scenarios) and the estimated bias seems to be minor.

Table 7.4: ERG EFS scenario analyses

EFS distribution	Treosulfan		Busulfan		Incr. Costs (£)	Incr. QALYs	ICER (£)
	Costs (£)	QALYs	Costs (£)	QALYs			
NMCM lognormal (base-case)	£128,147	6.49	£145,836	5.71	-£17,689	0.78	Treosulfan dominates
MCM lognormal	£127,787	6.49	£145,271	5.71	-£17,484	0.78	Treosulfan dominates
Gamma (treosulfan) NMCM lognormal (busulfan)	£131,678	6.47	£145,836	5.71	-£14,158	0.76	Treosulfan dominates

Based on electronic model
ERG = Evidence Review Group; ICER = incremental cost effectiveness ratio, EFS = event-free survival; Incr. = incremental, NMCM = non-mixture-cure model; QALY = quality-adjusted life year

7.2.2.2 Additional scenario 2: changing OS parametric distributions

The results of the scenarios assuming an MCM Weibull, an NMCM lognormal and an MCM lognormal distributions for OS are shown in Table 7.5. In all scenarios treosulfan dominated busulfan, with results similar to the base-case. The scenarios assuming NMCM/MCM lognormal distributions resulted in a larger gain in QALYs but less savings in costs.

Table 7.5: ERG OS scenario analyses

Scenario	Treosulfan		Busulfan		Incr. Costs (£)	Incr. QALYs	ICER (£)
	Costs (£)	QALYs	Costs (£)	QALYs			
NMCM Weibull (base-case)	£128,147	6.49	£145,836	5.71	-£17,689	0.78	Treosulfan dominates
MCM Weibull	£128,482	6.50	£146,658	5.72	-£18,177	0.78	Treosulfan dominates
NMCM lognormal	£127,813	6.49	£142,334	5.64	-£14,521	0.84	Treosulfan dominates
MCM lognormal	£127,735	6.48	£142,749	5.65	-£15,014	0.83	Treosulfan dominates

Based on electronic model
ERG = Evidence Review Group; ICER = incremental cost effectiveness ratio, Incr. = incremental, OS = overall survival, NMCM = non-mixture-cure model; QALY = quality-adjusted life year

7.2.2.3 Additional scenario 3: changing duration of cure point

The results obtained using different "cure points" are presented in Table 7.6. It can be seen that, only when the "cure point" was assumed to be one year, treosulfan did not dominate busulfan. In that scenario, the resulting ICER was £47,910 in the SW quadrant of the CE-plane; thus, treosulfan produced less QALYs than busulfan but also at lower costs. In the remaining scenarios, treosulfan always dominated. Assuming a "cure point" of three years resulted in the maximum incremental QALYs (1.03) and the minimum savings in costs (£4590). For "cure points" larger than five years, the value assumed in the base-case, costs savings increased while incremental QALYs remained approximately the same.

Table 7.6: ERG cure point scenario analyses

Scenario	Treosulfan		Busulfan		Incr. Costs (£)	Incr. QALYs	ICER (£)
	Costs (£)	QALYs	Costs (£)	QALYs			
1 year	£123,888	9.32	£125,181	9.35	-£1293	-0.03	£47,910*
2 years	£137,239	7.61	£144,629	6.84	-£7390	0.78	Treosulfan dominates
3 years	£132,395	6.88	£136,985	5.85	-£4590	1.03	Treosulfan dominates
4 years	£129,030	6.58	£140,595	5.68	-£11,565	0.90	Treosulfan dominates
5 years (ERG base-case)	£128,147	6.49	£145,836	5.71	-£17,689	0.78	Treosulfan dominates
6 years	£128,324	6.49	£150,148	5.77	-£21,823	0.72	Treosulfan dominates
7 years	£128,877	6.53	£153,606	5.84	-£24,728	0.69	Treosulfan dominates
8 years	£129,590	6.60	£156,631	5.92	-£27,041	0.69	Treosulfan dominates
9 years	£130,349	6.68	£159,367	5.99	-£29,017	0.69	Treosulfan dominates
10 years	£131,109	6.76	£161,893	6.07	-£30,784	0.70	Treosulfan dominates

Based on electronic model
* ICER in SW quadrant of the CE plane
ERG = Evidence Review Group; ICER = incremental cost effectiveness ratio, Incr. = incremental, QALY = quality-adjusted life year

7.2.2.4 Additional scenario 4: alternative approach to modelling long-term mortality (explicit vs. no explicit cure point)

The impact of selecting approach 3 to long-term mortality (use of parametric curves or HSCT-specific life tables – as determined by SMRs – depending on which has the highest mortality rate) instead of approach 5 (use of parametric curves up to the fixed "cure point", followed by switch to HSCT-specific

life table – as determined by SMRs – mortality rates) was explored by the ERG in this additional scenario. The results shown in Table 7.7 similar to the base-case results: treosulfan dominated but assuming approach 3 resulted in 0.03 more QALYs gained and £3,868 less savings in costs.

Table 7.7: ERG long-term mortality scenario analyses

Scenario	Treosulfan		Busulfan		Incr. Costs (£)	Incr. QALYs	ICER (£)
	Costs (£)	QALYs	Costs (£)	QALYs			
Base-case (approach 5)	£128,147	6.49	£145,836	5.71	-£17,689	0.78	Treosulfan dominates
Approach 3	£127,835	6.50	£141,656	5.69	-£13,821	0.81	Treosulfan dominates

Based on electronic model
ERG = Evidence Review Group; ICER = incremental cost effectiveness ratio, Incr. = incremental, QALY = quality-adjusted life year

7.2.2.5 Additional scenario 5: alternative SMRs to model HSCT mortality

The results obtained using different SMRs to calculate long-term post-HSCT mortality are presented in Table 7.8. It can be seen that in all scenarios treosulfan dominated busulfan. Assuming SMRs larger than in the base-case resulted in both smaller incremental QALYs and savings in costs. Assuming an SMR = 1 (i.e. mortality rate equal to UK general population) resulted in the maximum incremental QALYs (1.00) and the maximum savings in costs (£35,755). However, based on the feedback from the clinical experts (and confirmed by the company in the response to the request for clarification), this scenario should be considered as implausible.

Table 7.8: ERG SMRs for model HSCT mortality scenario analyses

Scenario	Treosulfan		Busulfan		Incr. Costs (£)	Incr. QALYs	ICER (£)
	Costs (£)	QALYs	Costs (£)	QALYs			
Mortality rate equal to UK general population (SMR = 1)	£145,584	7.63	£181,339	6.62	-£35,755	1.00	Treosulfan dominates
SMR for all alloHSCT patients (SMR = 4.30).	£126,415	5.54	£137,212	4.90	-£10,797	0.64	Treosulfan dominates
SMR for patients older than 45 years (SMR = 3.20).	£127,037	5.99	£140,503	5.29	-£13,466	0.70	Treosulfan dominates

SMR based on mean age + 5 years "cure point" (SMR = 2.88).	£127,329	6.15	£141,985	5.42	-£14,656	0.73	Treosulfan dominates
HR form NICE TA552[41] (HR = 2.30; base-case)	£128,147	6.49	£145,836	5.71	-£17,689	0.78	Treosulfan dominates
SMR for all HSCT patients (SMR = 4.50).	£126,339	5.48	£136,798	4.84	-£10,459	0.63	Treosulfan dominates
Based on electronic model ERG = Evidence Review Group; ICER = incremental cost effectiveness ratio, Incr. = incremental, QALY = quality-adjusted life year, TP = transition probability.							

7.2.2.6 Additional scenario 6: utilities

Several scenario analyses were run on the utility values utilised in the model. The results of these scenarios are presented in Table 7.9. In the first scenario, the Grulke values were still used for the health state utility values, but the Proskorovsky et al. 2014[75] mapping algorithm was replaced by the McKenzie et al. 2009[76] algorithm, as this was the algorithm that was considered the second best in this population. Use of the McKenzie et al. 2009[76] algorithm led to a decrease in QALYs gained in both treatment groups and a lower incremental QALY, but treosulfan still dominates. The second scenario utilised the DCE values available for health state utility values from Castejon et al. 2018[59] and for adverse events from Stein et al. 2018[62]. This scenario led to lower QALYs gained in both treatment groups but higher incremental QALYs than the ERG base-case. Again, treosulfan still dominates. The third scenario uses the base-case utility values, except for utilising the MDS specific utility value from Szende et al. 2009.[64] This has very little impact on the QALYs gained and treosulfan still dominates. The final scenario uses the utility values from TA399[43], using the Proskorovsky et al. 2014[75] mapping algorithm. This results in higher QALYs gained in both treatment groups but no noticeable difference in incremental QALYs and so treosulfan still dominates.

Table 7.9: ERG utility scenario analyses

Scenario	Treosulfan		Busulfan		Incr. Costs (£)	Incr. QALYs	ICER (£)
	Costs (£)	QALYs	Costs (£)	QALYs			
ERG preferred base-case	£128,147	6.49	£145,836	5.71	-£17,689	0.78	Treosulfan dominates

Scenario	Treosulfan		Busulfan		Incr. Costs (£)	Incr. QALYs	ICER (£)
	Costs (£)	QALYs	Costs (£)	QALYs			
McKenzie mapping[76] of Grulke values [71]	£128,147	5.53	£145,836	4.84	-£17,689	0.69	Treosulfan dominates
DCE scenario	£128,147	5.45	£145,836	4.63	-£17,689	0.82	Treosulfan dominates
MDS specific relapse utility	£128,147	6.49	£145,836	5.70	-£17,689	0.79	Treosulfan dominates
TA399 utility values[43]	£128,147	6.83	£145,836	6.05	-£17,689	0.78	Treosulfan dominates
Based on electronic model DCE = discrete choice experiment; ERG = Evidence Review Group; ICER = incremental cost effectiveness ratio, Incr. = incremental, QALY = quality-adjusted life year							

7.2.2.7 Additional scenario 7: resource use and costs

In all three cost scenario analyses that were conducted (both 100% and 0%, instead of 50% in base-case, of relapsed patients receive a second HSCT during late relapse in cost scenarios 1 and 2, respectively, and in cost scenario 3 0% wastage, instead of 100% in base-case) of treosulfan administration is assumed), the influence of the changed assumptions on the results are minor, and none of them leads to different conclusions. The results are shown below in Table 7.10.

Table 7.10: ERG costs and resource use scenario analyses

Scenario	Treosulfan		Busulfan		Incr. Costs (£)	Incr. QALYs	ICER (£)
	Costs (£)	QALYs	Costs (£)	QALYs			
ERG preferred base-case	£128,147	6.49	£145,836	5.71	-£17,689	0.78	Treosulfan dominates
Scenario 1	£129,273	6.49	£145,690	5.71	-£16,416	0.78	Treosulfan dominates
Scenario 2	£127,021	6.49	£145,982	5.71	-£18,961	0.78	Treosulfan dominates
Scenario 3	£128,060	6.49	£145,834	5.71	-£17,774	0.78	Treosulfan dominates
Based on electronic model ERG = Evidence Review Group; ICER = incremental cost effectiveness ratio, Incr. = incremental, QALY = quality-adjusted life year							

7.2.2.8 Additional scenario 8: subgroup analyses

Based on the analyses presented in Appendix 2, the ERG modelled OS according to the MCM lognormal distribution and EFS according to the NMCM lognormal distribution for the AML subpopulation. The results for this subgroup are presented in Table 7.11 and indicated that treosulfan

was dominant. Compared to the company results in Table 6.5, both cost savings and incremental QALYs for treosulfan were larger in the company's analysis.

Table 7.11: ERG AML subgroup cost effectiveness results

Technologies	Total costs (£)	Total LYGs	Total QALYs	Incr. costs (£)	Incr. LYGs	Incr. QALYs	ICER versus baseline (£/QALY)
Treosulfan	£131,425	8.96	6.63	-£24,949	0.75	0.71	Treosulfan dominates
Busulfan	£156,374	8.21	5.93				
Based on electronic model AML = acute myeloid leukaemia; CS = company submission; ICER = incremental cost effectiveness ratio; Incr. = incremental; LYGs = life years gained; QALYs = quality-adjusted life years							

The results of the scenario where a Gamma distribution for EFS was assumed (Table 7.12) were similar to those presented above.

Table 7.12: ERG AML subgroup cost effectiveness results (alternative scenario)

Technologies	Total costs (£)	Total LYGs	Total QALYs	Incr. costs (£)	Incr. LYGs	Incr. QALYs	ICER versus baseline (£/QALY)
Treosulfan	£136,509	8.96	6.60	-£31,139	0.75	0.74	Treosulfan dominates
Busulfan	£167,648	8.21	5.87				
Based on electronic model AML = acute myeloid leukaemia; CS = company submission; ICER = incremental cost effectiveness ratio; Incr. = incremental; LYGs = life years gained; QALYs = quality-adjusted life years							

For the MDS subpopulation, the ERG modelled OS according to a NMCM Weibull distribution and EFS according to a MCM lognormal distribution. The results for the MDS subgroup are shown in Table 7.13 and also indicated that treosulfan was dominant. Compared to the company results in Table 6.6, cost savings were smaller and incremental QALYs for treosulfan were larger in the company's analysis.

Table 7.13: ERG MDS subgroup cost effectiveness results

Technologies	Total costs (£)	Total LYGs	Total QALYs	Incr. costs (£)	Incr. LYGs	Incr. QALYs	ICER versus baseline (£/QALY)
Treosulfan	£120,438	8.19	6.11	-£6257	1.06	0.85	Treosulfan dominates
Busulfan	£126,695	7.14	5.26				
Based on electronic model CS = company submission; ICER = incremental cost effectiveness ratio; Incr. = incremental; LYGs = life years gained; MDS = myelodysplastic syndrome; QALYs = quality-adjusted life years.							

Since the choice of EFS survival curves made above seemed to overestimate survival gains for treosulfan, the results of an additional scenario assuming a lognormal distribution for treosulfan and an MCM Weibull for busulfan are presented in Table 7.14. The results indicated that, compared with busulfan, treosulfan generated 0.79 incremental QALYs, and 1.06 incremental LYGs, with (higher) incremental costs of £6456. Thus, despite not being dominant, treosulfan can be considered cost effective since the ICER was £8,171 per QALY gained, well below the common £20,000 or £30,000 willingness to pay thresholds.

Table 7.14: ERG MDS subgroup cost effectiveness results (alternative scenario)

Technologies	Total costs (£)	Total LYGs	Total QALYs	Incr. costs (£)	Incr. LYGs	Incr. QALYs	ICER versus baseline (£/QALY)
Treosulfan	£130,840	8.19	6.06	£6456	1.06	0.79	£8171
Busulfan	£124,384	7.14	5.27				

Based on electronic model
CS = company submission; ICER = incremental cost effectiveness ratio; Incr. = incremental; LYGs = life years gained; MDS = myelodysplastic syndrome; QALYs = quality-adjusted life years

In conclusion, in both subgroups treosulfan was dominant or cost effective. In the AML subgroup, cost savings for treosulfan were larger than in the MDS subgroup, but with smaller QALY gains. The uncertainty in the MDS subgroup seems to be larger too, given the overall poor fit of the OS and specially EFS curves. A PSA for the MDS subgroup only might be informative. Unfortunately, the ERG was unable to run such an analysis with the current version of the electronic model.

7.3 ERG's preferred assumptions

The ERG preferred changes to the updated company base-case were described in Section 7.1.2 of this report. The cost effectiveness results of the ERG preferred base-case are presented in Table 7.15 in four steps. In each step, the cumulative impact on the model results is shown. The assumption with the largest impact on the incremental costs was correcting the calculations of the overall and event-free survival probabilities. This results in incremental costs increased by £9,176 and incremental QALYs decreased by 0.05. The assumption with the largest impact on the incremental QALYs was modelling OS according to a NMCM Weibull distribution. This results in incremental QALYs decreased by 0.06 and incremental costs decreased by £3,151. The other two changes made by the ERG had a minor impact on the results.

Table 7.15: ERG's preferred model assumptions

Preferred assumption	Section in ERG report	Treosulfan		Busulfan		Inc. Costs (£)	Inc. QALYs	Cumulative ICER (£/QALY)
		Total Costs (£)	Total QALYs	Total Costs (£)	Total QALYs			
Company base-case	6.1	£137,062	6.44	£160,821	5.55	-£23,759	0.89	Treosulfan dominates
Company updated base-case (after clarification)	7.1.1	£140,308	6.44	£163,976	5.55	-£23,668	0.89	Treosulfan dominates
ERG change 1 – Correct OS and EFS implementation	7.1.2	£127,823	6.50	£142,315	5.65	-£14,492	0.84	Treosulfan dominates
ERG change 2 – Rescaling factor (year to day)	7.1.2	£127,807	6.48	£142,297	5.64	-£14,490	0.84	Treosulfan dominates
ERG change 3 – Modelling OS according to a NMCM Weibull distribution	5.2.6.1	£128,142	6.48	£145,783	5.70	-£17,641	0.78	Treosulfan dominates
ERG change 4 – Most recent life tables (with correct formula)	7.1.2	£128,147	6.49	£145,836	5.71	-£17,689	0.78	Treosulfan dominates

CS = company submission; ERG = Evidence Review Group; ICER = incremental cost effectiveness ratio; QALY = quality adjusted life year

7.4 Conclusions of the cost effectiveness section

The company presented a *de novo* partitioned-survival model to evaluate the cost effectiveness of treosulfan with fludarabine compared to busulfan with fludarabine as a conditioning treatment for malignant disease (AML or MDS) prior to allogeneic haematopoietic stem cell transplantation in adults. The simulation assumes that all patients have received an HSCT and all patients start in the remission health state. From there, patients can either relapse or die. Transitions between health states are determined by EFS and OS curves obtained by fitting statistical models to the survival data from the MC-FludT.14/L Trial II. Health states costs and utilities are used to calculate total costs and total QALYs over a lifetime time horizon.

The company used data from MC-FludT.14/L Trial II to inform OS and EFS in the model. Standard parametric models (e.g. exponential, Weibull, Gompertz, etc.) and MCMs/NMCMs were considered by the company.

Two main aspects were considered by the company when modelling long-term mortality: the selection of a "cure point" and the application of an SMR for background mortality. The rationale for including a "cure point" in the model is that HSCT is a potentially curative treatment. Prior to this "cure point", transitions between the health states of the model are calculated based on the parametric OS and EFS curves. After the cure point, only background mortality is determining the transitions between the health states. Based on the opinion of two clinical experts, the company assumed a fixed "cure point" of five years. This means that patients surviving alloHSCT for at least five years are considered as "cured" so that relapses/transplant-related deaths after five years are very rare. The HR = 2.30 used in the TA552[41] was assumed to calculate long-term mortality of patients post-HSCT, since this was the value deemed as the most appropriate by the clinical experts consulted by the company. Besides the selection of the "cure point" and the SMR for background mortality, the company also included in the model five approaches to long-term mortality, from which only two were plausible: 1) using parametric curves or HSCT-specific life tables (as determined by SMRs), depending on which has the highest mortality rate, and 2) using parametric curves up to the fixed "cure point", followed by switch to HSCT-specific life table (as determined by SMRs) mortality rates. The latter was assumed by the company for the base-case, since this was deemed to be the most appropriate and reflective of AML and MDS patients by the clinical experts consulted by the company.

All treatment related Grade 3+ Common Terminology Criteria for Adverse Events (CTCAEs) with an incidence ≥1% in the trial were taken into account, including extensive cGvHD and stage III/IV aGvHD. Durations for extensive cGvHD and stage III/IV aGvHD were based on TA545[44] and clinical expert opinion, weighted averages of elective inpatient codes for items costed using NHS reference costs,[51] and trial estimates for those items costed using NHS Tariffs.[95]

HRQoL data were not collected in the trial. The company conducted a systematic search of the literature and an additional targeted search of the literature, including relevant NICE technology appraisals. The company identified a range of sources of utility evidence; however, none met the NICE reference case preference for UK EQ-5D values. Therefore, in the base-case the company utilised health state utility values from Grulke et al. 2012,[71] mapping from the QLQ-C30 to the UK EQ-5D utility values using an established and appropriate mapping algorithm developed by Proskorovsky et al. 2014.[75] Health state utilities for relapse/progression patients were estimated using a disutility multiplier estimated from TA399 based on mappings of QLQ-C30 data to EQ-5D using Proskorovsky et al. 2014,[75] which was applied to the short-term post-HSCT recovery (≤6 months) estimated from Grulke et al. 2012.[71] Base-

case disutilities for GvHD were identified from Kurosawa et al. 2016[63] based on non-UK EQ-5D values. Disutilities for Grade 3 and 4 AEs were identified from TA399[43] based on mappings from QLQ-C30 data, which were mapped to UK EQ-5D values, again using the Proskorovsky et al. 2014 mapping algorithm.[75] Utilities applied in the model were adjusted for age and gender using the methods of Ara and Brazier 2010.[77]

Treatment costs included costs for intravenous treosulfan and busulfan (both based on trial regimens), and concomitant medication. Post-HSCT recovery costs in EFS for the first 24 months were based on an NHS Blood and Transplant Analysis 2015.[89] After 24 months, these costs were based on clinical expert opinion. Resource use in the relapsed/progressed disease health state included monitoring costs and blood transfusion costs (both sourced from TA399[43]), and treatment costs (based on clinical expert opinion). A mortality, sourced from PSSRU 2017/2018,[42] was also included in the model. Adverse event costs for extensive cGvHD were sourced from Robin et al. 2017[92] and stage III/IV aGvHD costs, from NHS reference costs if available,[51] and from NHS tariffs otherwise.[95]

The discounted base-case results indicated that treosulfan generated 0.89 more QALYs than busulfan at lower costs (-£23,759). Therefore, treosulfan was dominant over busulfan. The disaggregated results by health state showed that, compared to busulfan, treosulfan approximately halved the time in the relapse/progression health state. Since the costs of the relapse/progression health state are high, the relapse/progression costs saved by treosulfan outweighed the lower busulfan treatment and EFS costs. Hence, the dominance. The results from the PSA were in line with the deterministic ones since treosulfan was also dominant over busulfan. The vast majority of the 5,000 iterations provided results in the eastern quadrants of the CE plane. In particular, 69.6% of the iterations were in the south-eastern quadrant of the CE plane. The CEAC also showed that treosulfan has an 84.3% probability of being cost effective at a threshold of £20,000 per QALY and an 89.9% probability of being cost effective at a threshold of £30,000 per QALY. The company also conducted a one-way deterministic sensitivity analysis. The results of this analysis indicated that only changes in the meanlog coefficients of the busulfan EFS distribution and the treosulfan OS distribution resulted in a negative INMB or, equivalently, in an ICER above £30,000 per QALY. All the scenario analyses conducted by the company resulted in treosulfan dominating busulfan. Even in the scenario that was the least favourable for treosulfan, when a five-year time horizon was assumed, treosulfan still generated more QALYs than busulfan (0.29) at lower costs (-£6,285). Finally, the company performed subgroup analysis for AML and MDS patients separately. In both subgroups treosulfan was dominant. In the AML subgroup, cost savings for treosulfan were larger than in the base-case, but with smaller QALY gains. The opposite was observed for the MDS subgroup.

The modelling approach considered by the company is in line with those in two recent NICE technology appraisals on AML (TA523[40] and TA552[41]). The ERG considers this approach appropriate for the decision problem at hand. Even though graft failure was included in the definition of EFS in MC-FludT.14/L Trial II, the company included graft failure as an adverse "event" in terms of EFS and not as a health state of the model. This choice was made based on the low number of graft failure events observed in the trial (which would make extrapolating long-term survival curves for graft failure patients unreliable) and because the company did not identify any data indicating differences between relapse/progression patients and graft failure patients in terms of either costs or quality of life. The ERG agrees with this approach.

Despite indicating AIC, BIC, visual inspection and KOL feedback as specific criteria for assessing goodness of fit, assessments based on visual inspection and KOL feedback were not reported in the CS.

For that reason, the ERG conducted a more detailed overall goodness-of-fit assessment based on all the information presented by the company either in the main submission document, in the appendices or in the response to the request for clarification. Based on this assessment, the ERG was able to identify the best candidate distributions to model OS and EFS, which in some cases did not match the ones selected by the company, and to define alternative scenario analyses.

Justification for adverse event, resource use, costs and utility sources chosen by the company was clear and the ERG agreed with the company's choice in relation to the application of costs and utilities in the model. Therefore, the ERG did not amend the costs or utility parameters chosen by the company for the base-case. The robustness of the model results to changes in both cost and utility parameters was tested with scenario analyses, conducted by both the company and the ERG.

Following the clarification questions from the ERG, the company corrected negative costs appearing at the end of the model time horizon and amended the age decrement in utility associated to adverse events. The effect of these changes on the base-case results was minor. Additionally, the ERG corrected the implementation of OS end EFS in the model as explained in Appendix 2, and used a factor 1/365.25 (instead of 1/364) to re-scaled yearly values into to daily. Furthermore, the ERG modelled OS according to a NMCM Weibull distribution and used the most recent version of the UK life tables. The results of the ERG preferred base-case analysis resulted in treosulfan generating 0.78 more QALYs than busulfan at lower costs (-£17,689). Therefore, treosulfan dominated busulfan as in the company base-case. In the company base-case, treosulfan generated 0.89 more QALYs than busulfan and saved -£23,759. Thus, in the ERG base-case, both cost savings and QALY gains for treosulfan were smaller than in the company base-case.

The ERG also conducted a PSA using the ERG preferred base-case assumptions. The probabilistic results were in line with the deterministic ones. In particular, 72.1% of the iterations were in the south-eastern quadrant of the CE plane, where treosulfan dominates busulfan. The CEAC also showed that treosulfan has an 89.9% probability of being cost effective at a threshold of £20,000 per QALY and a 94.1% probability of being cost effective at a threshold of £30,000 per QALY. These results are in line with those presented in the company base-case. The ERG did not conduct a one-way sensitivity analysis. The one-way sensitivity analysis was conducted, allowing 20% variation from mean value for all parameters. This seems arbitrary to the ERG and may not represent an equally plausible range of variation for all input parameters. Similar to the PSA described in the CS, the limits of (95%) confidence intervals should be used for each parameter to calculate the upper and lower bounds of the bars in the tornado diagram. However, the ERG acknowledges that the choice of consistent parameter variation for the deterministic sensitivity analysis is in line with several recent NICE technology appraisals for AML (TA399, TA523, TA545), and that covariance between input parameters, particularly when a large number of input variables are included, may affect the ability to appropriately interpret results of deterministic sensitivity analyses. Furthermore, the NICE reference case does not provide explicitly statements on what upper and lower limits should be adopted for deterministic sensitivity analysis. Given the time constraints associated with this project, the ERG was not able to test the impact of using 95% confidence intervals for the deterministic sensitivity analysis.

Based on the base-case, sensitivity and scenario analyses presented by the company, the ERG considers that the results of the model are very robust and that any additional scenarios are likely to result in treosulfan dominance. Since the costs associated to the relapse health state are high, even a minor reduction in the time spent in the relapse health state will be sufficient for treosulfan dominance. Despite this, the ERG considered that potentially important scenarios to test the main structural uncertainties in

the economic model were missing in the CS. Therefore, several additional scenarios were conducted by the ERG. Assumptions regarding the selection of parametric curves for EFS and OS, the duration of the "cure point", alternative approaches to long-term mortality or different SMRs to model HSCT mortality should have been explored by the company. The results of these analyses confirmed the ERG expectations since in all the scenarios explored, treosulfan was dominant. Finally, the ERG also presented results for AML and MDS patients separately. In these analyses, the company assumed the same EFS and OS distributions as in the base-case with the pooled population. However, the ERG considered that the selection of survival models should be based on subgroup-specific data. Based on these analyses, the ERG concluded that for the AML subpopulation the preferred OS distribution was the MCM lognormal and the preferred EFS distribution was the NMCM lognormal. For the MDS subpopulation the preferred OS distribution was the NMCM Weibull and the preferred EFS distribution was the MCM lognormal. Furthermore, all subgroup analyses were run selecting subgroup-specific patient characteristics (instead of using the characteristics of the pooled population). In both subgroups treosulfan was dominant or cost effective. In the AML subgroup, cost savings for treosulfan were larger than in the MDS subgroup, but with smaller QALY gains. The uncertainty in the MDS subgroup seems to be larger too, given the overall poor fit of the OS and specially EFS curves. A PSA for the MDS subgroup only might be informative. Unfortunately, the ERG was unable to run such an analysis with the current version of the electronic model.

8. End of life

The company submission did not include any statement indicating that end of life criteria apply.

9. References

[1] Medac. *Treosulfan with fludarabine for malignant disease before allogeneic stem cell transplant [ID1508]: submission to National Institute for Health and Care Excellence. Single technology appraisal (STA)*. Wedel: Medac, 2019. 211p.

[2] Majhail NS, Farnia SH, Carpenter PA, Champlin RE, Crawford S, Marks DI, et al. Indications for Autologous and Allogeneic Hematopoietic Cell Transplantation: Guidelines from the American Society for Blood and Marrow Transplantation. *Biol Blood Marrow Transplant* 2015;21(11):1863-1869.

[3] Abdul Wahid SF, Ismail NA, Mohd-Idris MR, Jamaluddin FW, Tumian N, Sze-Wei EY, et al. Comparison of reduced-intensity and myeloablative conditioning regimens for allogeneic hematopoietic stem cell transplantation in patients with acute myeloid leukemia and acute lymphoblastic leukemia: a meta-analysis. *Stem Cells Dev* 2014;23(21):2535-52.

[4] British Society of Blood and Marrow Transplantation. BSBMT Registry 2017 [Internet]. London: BSBMT, 2019 [accessed 12.7.19]. Available from: http://bsbmt.org/activity/2017/

[5] Carreras E DC, Mohty M, Kröger, editors. *The EBMT handbook: hematopoietic stem cell transplantation and cellular therapies [Internet]*. London: EBMT, 2019 [accessed 12.7.19]. 688p. Available from: https://www.ebmt.org/sites/default/files/2019-01/2019_Book_TheEBMTHandbook.pdf

[6] Scheulen ME, Hilger RA, Oberhoff C, Casper J, Freund M, Josten KM, et al. Clinical phase I dose escalation and pharmacokinetic study of high-dose chemotherapy with treosulfan and autologous peripheral blood stem cell transplantation in patients with advanced malignancies. *Cancer Res* 2000;6:4209-16.

[7] Sehouli J, Tome O, Dimitrova D, Camara O, Runnebaum IB, Tessen HW, et al. A phase III, open label, randomized multicenter controlled trial of oral versus intravenous treosulfan in heavily pretreated recurrent ovarian cancer: a study of the North-Eastern German Society of Gynecological Oncology (NOGGO). *J Cancer Res Clin Oncol* 2017;143(3):541-550.

[8] National Institute for Health and Care Excellence. *Single technology appraisal. Treosulfan with fludarabine for malignant disease before allogeneic stem cell transplant: final scope (Appendix B) [Internet]*. London: NICE, 2019 [accessed 9.7.19]. 4p. Available from: https://www.nice.org.uk/guidance/gid-ta10421/documents/final-scope

[9] European Medicines Agency. Trecondi [Internet]. Amsterdam: EMA, 2019 [accesed 25.7.19]. Available from: https://www.ema.europa.eu/en/medicines/human/EPAR/trecondi

[10] Medac. *Clinical phase III trial to compare treosulfan-based conditioning therapy with busulfan-based reduced-intensity conditioning (RIC) prior to allogeneic haematopoietic stem cell transplantation in patients with AML or MDS considered ineligible to standard conditioning regimens*

– final analysis including post-surveillance. (Clinical trial report: treosulfan) Protocol MC-FludT.14/L - Part II [PDF supplied with the Company's submission]: Medac, 25 Jan 2018. 1599p.

[11] National Institute for Health and Care Excellence. *Treosulfan with fludarabine for malignant disease before allogeneic stem cell transplant [ID1508]: Clarification letter*. London: NICE, 2019. 6p.

[12] Medac. *Treosulfan with fludarabine for malignant disease before allogeneic stem cell transplant [ID1508] – Response to request for clarification from the ERG*: Medac, 2019. 85p.

[13] Medac. *Annex 1: summary of product characteristics. Trecondi 1 g powder for solution for infusion; trecondi 5 g powder for solution for infusion*. Wedel: Medac, [2018]. 37p.

[14] EMC. Treosulfan injection (medac UK) [Internet]. Leatherhead: Datapharm, 2017 [accessed 23.7.19]. Available from: https://www.medicines.org.uk/emc/product/1414/smpc

[15] Iacobelli S, Eikema D-J, Koster L, van Biezen A, European society for Blood and Marrow Transplantation. *Re-analysis of EBMT-registry data on fludarabine/melphalan and busulfan/cyclophosphamide based conditioning treatment compared to fludarabine/treosulfan based conditioning of adult AML and MDS patients treated in MC-FludT.14/L phase III by matched pairs. Statistical analysis report: 2010-2016*: EBMT, 18 Mar 2019. 190p.

[16] Center for International Blood & Marrow Transplant Research. *Comparison of allogeneic transplantation outcomes for acute myeloid leukemia and myelodysplastic syndrome with full or less intensity busulfan-fludarabine regimens and busulfan-cyclophosphamide regimen in the United States (US) to a treosulfan-containing regimen in a medac-sponsored European (EU) clinical trial MC-FludT.14/L (Trial II)*. Milwaukee (WI): CEBMTR, Feb 2019. 27p.

[17] Medac. *Treosulfan with fludarabine for malignant disease before allogeneic stem cell transplant [ID1508]: Appendix D - Identification, selection and synthesis of clinical evidence*. Wedel: Medac, 2019. 290p.

[18] Medac. *Treosulfan with fludarabine for malignant disease before allogeneic stem cell transplant [ID1508]: Appendix L - Additional information company evidence submission*. Wedel: Medac, 2019. 101p.

[19] Passweg JR, Baldomero H, Bader P, Basak GW, Bonini C, Duarte R, et al. Is the use of unrelated donor transplantation leveling off in Europe? The 2016 European Society for Blood and Marrow Transplant activity survey report. *Bone Marrow Transplant* 2018;53(9):1139-1148.

[20] Champlin R. Reduced intensity allogeneic hematopoietic transplantation is an established standard of care for treatment of older patients with acute myeloid leukemia. *Best Pract Res Clin Haematol* 2013;26(3):297-300.

[21] Dhawan R, Marks DI. Who should receive a transplant for acute lymphoblastic leukaemia? *Curr Hematol Malig Rep* 2017;12(2):143-52.

[22] Oliansky DM, Larson RA, Weisdorf D, Dillon H, Ratko TA, Wall D, et al. The role of cytotoxic therapy with hematopoietic stem cell transplantation in the treatment of adult acute lymphoblastic leukemia: update of the 2006 evidence-based review. *Biol Blood Marrow Transplant* 2012;18(1):18-36 e6.

[23] National Institute for Health and Care Excellence. *Methods for the development of NICE public health guidance (third edition): process and methods [Internet]*. London: NICE, 2012 [accessed 3.7.19]. 269p. Available from: nice.org.uk/process/pmg4

[24] Higgins JP, Altman DG, Gotzsche PC, Juni P, Moher D, Oxman AD, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ* 2011;343:d5928.

[25] Higgins JP, Savović J, Page MJ, Sterne JA. *Revised Cochrane risk-of-bias tool for randomized trials (RoB 2) [Internet]*: ROB2 Development Group, 2019 [accessed 12.7.19] Available from: https://drive.google.com/open?id=1xBzD1BeRrXH2DHg58fnTNKinTkX5SLEt

[26] Medac Research. *Allogeneic hematopoietic stem cell transplantation in patients with advanced hematological malignancies after treosulfan-based conditioning therapy – a clinical phase II study. (Final (clinical) study report: treosulfan) Protocol MC-FludT.6/L [PDF supplied with the Company's submission]*. Hamburg: Medac, 13 Aug 2009. 229p.

[27] Medac Research. *Clinical phase II trial to evaluate the safety and efficacy of treosulfan based conditioning prior to allogeneic haematopoietic stem cell transplantation in patients with acute myeloid leukaemia. (Final study report: treosulfan) Protocol MC-Flud T.7/AML [PDF supplied with the Company's submission]*. Hamburg: Medac, 7 Jul 2011. 194p.

[28] Medac Research. *Clinical phase II trial to evaluate the safety and efficacy of treosulfan based conditioning prior to allogeneic haematopoietic stem cell transplantation in patients with myelo-dysplastic syndrome (MDS). (Final study report: treosulfan) Protocol MC-FludT.8/MDS [PDF supplied with the Company's submission]*. Hamburg: Medac, 9 Dec 2010. 200p.

[29] Medac. *Clinical Phase 2 trial to describe the safety and efficacy of Treosulfan-based conditioning therapy prior to allogeneic haematopoietic stem cell transplantation in paediatric patients with haematological malignancies. (Clinical trial report: treosulfan) Protocol MC-FludT.17/M [PDF supplied with the Company's submission]*. Wedel: Medac, 12 Mar 2018. 1028p.

[30] National Health Service. Stem cell and bone marrow transplants [Internet]. NHS, 2018 [accessed 3.7.19]. Available from: https://www.nhs.uk/conditions/stem-cell-transplant/

[31] Swerdlow SH CE, Harris NL, Jaffe ES, Pileri SA, Stein H, Thiele J, Vardiman JW, editors. *WHO classification of tumours of haematopoietic and lymphoid tissues [Internet]*. Lyon: International Agency for Research on Cancer, 2008 [accessed 10.7.19]. 444p. Available from: www.pathologyoutlines.com/site/BBHeme.pdf

[32] Sorror ML, Maris MB, Storb R, Baron F, Sandmaier BM, Maloney DG, et al. Hematopoietic cell transplantation (HCT)-specific comorbidity index: a new tool for risk assessment before allogeneic HCT. *Blood* 2005;106(8):2912-9.

[33] Yerushalmi R, Shem-Tov N, Danylesko I, Avigdor A, Nagler A, Shimoni A. Fludarabine and treosulfan compared with other reduced-intensity conditioning regimens for allogeneic stem cell transplantation in patients with lymphoid malignancies. *Bone Marrow Transplant* 2015;50(12):1526-35.

[34] Shimoni A, Galski H, Iacobelli S, van Biezen A, Beelen D, Mufti GJ, et al. Fludarabine and treosulfan conditioning is associated with a more favorable outcome after allogeneic stem cell transplantation in myelodysplastic syndrome. A survey on behalf of the Chronic Malignancies Working Party of EBMT. Presented at 41st Annual Meeting of the EBMT; 22-25 March 2015; Istanbul. 2015.

[35] Shimoni A, Labopin M, Savani B, Hamladji RM, Beelen D, Mufti G, et al. Intravenous busulfan compared with treosulfan-based conditioning for allogeneic stem cell transplantation in acute myeloid leukemia: a study on behalf of the Acute Leukemia Working Party of European Society for Blood and Marrow Transplantation. *Biol Blood Marrow Transplant* 2018;24(4):751-7.

[36] Saraceni F, Labopin M, Brecht A, Kroger N, Eder M, Tischer J, et al. Fludarabine-treosulfan compared to thiotepa-busulfan-fludarabine or FLAMSA as conditioning regimen for patients with primary refractory or relapsed acute myeloid leukemia: a study from the Acute Leukemia Working Party of the European Society for Blood and Marrow Transplantation (EBMT). *J Hematol Oncol* 2019;12(1):44.

[37] Gran C, Wang J, Nahi H, Koster L, Gahrton G, Wolschke C, et al. Treosulfan conditioning for allogeneic transplantation in multiple myeloma improved overall survival in upfront hematopoietic stem cell transplantation — a large retrospective study by the Chronic Malignancies Working Party of the EBMT. *Blood* 2018;132(Suppl 1):3464-3464.

[38] Medac. *MC-FludT.14/L: statistical analysis plan (treosulfan) [PDF supplied with the Company's submission]*. Wedel: Medac, 2018. 46p.

[39] Higgins JPT, Green S, eds. *Cochrane handbook for systematic reviews of interventions [Internet]*. Version 5.1.0 [updated March 2011]: The Cochrane Collaboration, 2011 [accessed 23.3.11]. Available from: http://handbook.cochrane.org/

[40] National Institute for Health and Care Excellence. *Midostaurin for untreated acute myeloid leukaemia: NICE technology appraisal guidance 523 (TA523) [Internet]*. London: NICE, 2018 [accessed 12.7.19]. 21p. Available from: www.nice.org.uk/guidance/ta523

[41] National Institute for Health and Care Excellence. *Liposomal cytarabine-daunorubicin for untreated acute myeloid leukaemia: NICE technology appraisal guidance 552 (TA552) [Internet]*. London: NICE, 2018 [accessed 12.7.19]. 18p. Available from: www.nice.org.uk/guidance/ta552

[42] Curtis LA, Burns A, Personal Social Services Research Unit. *Unit Costs of Health and Social Care 2018 [Internet]*. Canterbury: University of Kent, 2018 [accessed 25.7.19]. 203p. Available from: https://www.pssru.ac.uk/project-pages/unit-costs/

[43] National Institute for Health and Care Excellence. *Azacitidine for treating acute myeloid leukaemia with more than 30% bone marrow blasts: NICE technology appraisal guidance 399 (TA399) [Internet]*. London: NICE, 2016 [accessed 12.7.19]. 36p. Available from: www.nice.org.uk/guidance/ta399

[44] National Institute for Health and Care Excellence. *Gemtuzumab ozogamicin for untreated acute myeloid leukaemia: NiCE technology appraisal guidance 545 (TA545) [Internet]*. London: NICE, 2018 [accessed 12.7.19]. 24p. Available from: www.nice.org.uk/guidance/ta545

[45] National Institute for Health and Care Excellence. *Guide to the methods of technology appraisal 2013 [Internet]*. London: NICE, 2013 [accessed ????]. 93p. Available from: http://publications.nice.org.uk/pmg9

[46] Latimer N. *NICE DSU Technical Support Document 14: Survival analysis for economic evaluations alongside clinical trials - extrapolation with patient-level data. Report by the Decision Support Unit [Internet]*. Sheffield: ScHARR, 2013 [accessed 23.7.19]. 52p. Available from: http://nicedsu.org.uk/wp-content/uploads/2016/03/NICE-DSU-TSD-Survival-analysis.updated-March-2013.v2.pdf

[47] Mamolo CM, Cappelleri JC, Hoang CJ, Kim R, Hadfield A, Middleton C, et al. Cross-sectional survey of symptoms and health-related quality of life of adults with de novo acute myeloid leukemia (AML) in clinical practice. *Blood* 2017;130(5660):1-3.

[48] Kurosawa S, Yamaguchi H, Yamaguchi T, Fukunaga K, Yui S, Kanamori H, et al. Decision analysis of allogeneic hematopoietic stem cell transplantation versus chemotherapy in cytogenetically standard-risk acute myeloid leukemia in first complete remission: the impact of FLT3-ITD profile. *Blood* 2014;124(21):1221-1221.

[49] Martin PJ, Counts GW, Jr., Appelbaum FR, Lee SJ, Sanders JE, Deeg HJ, et al. Life expectancy in patients surviving more than 5 years after hematopoietic cell transplantation. *J Clin Oncol* 2010;28(6):1011-6.

[50] Medac. *Treosulfan with fludarabine for malignant disease before allogeneic stem cell transplant [ID1508]. Appendix I - cost and healthcare resource identification, measurement and valuation.* Wedel: Medac, 2019. 24p.

[51] NHS Improvement. NHS reference costs 2017/18 [Internet]. London: NHS Improvement, 2018 [accessed 1.4.19]. Available from: https://improvement.nhs.uk/resources/reference-costs/

[52] Kurosawa S, Yamaguchi T, Miyawaki S, Uchida N, Kanamori H, Usuki K, et al. A Markov decision analysis of allogeneic hematopoietic cell transplantation versus chemotherapy in patients with acute myeloid leukemia in first remission. *Blood* 2011;117(7):2113-20.

[53] Stauder R, Yu G, Koinig K, Fenaux P, Symeonidis A, Sanz G, et al. Health-Related Quality of Life is Substantially Impaired in Lower-Risk MDS when Compared with Reference Populations and Significantly Affects Overall Survival. *Leuk Res* 2017;55:S12.

[54] Stauder R, Yu G, Koinig KA, Bagguley T, Fenaux P, Symeonidis A, et al. Health-related quality of life in lower-risk MDS patients compared with age- and sex-matched reference populations: a European LeukemiaNet study. *Leukemia* 2018;32(6):1380-1392.

[55] Kayastha N, Wolf SP, Locke SC, Samsa GP, El-Jawahri A, LeBlanc TW. The impact of remission status on patients' experiences with acute myeloid leukemia (AML): an exploratory analysis of longitudinal patient-reported outcomes data. *Support Care Cancer* 2018;26(5):1437-1445.

[56] Buckley SA, Jimenez-Sahagun D, Othus M, Walter RB, Lee SJ. Quality of life from the perspective of the patient with acute myeloid leukemia. *Cancer* 2018;124(1):145-152.

[57] Ramos F, Pedro C, Tormo M, de Paz R, Font P, Luno E, et al. Impact of anaemia on health-related quality of life and cardiac remodelling in patients with lower risk myelodysplastic syndromes. Results of GlobQoL study. *Eur J Cancer Care (Engl)* 2017;26(6).

[58] Messerer D, Engel J, Hasford J, Schaich M, Ehninger G, Sauerland C, et al. Impact of different post-remission strategies on quality of life in patients with acute myeloid leukemia. *Haematologica* 2008;93(6):826-33.

[59] Castejon N, Cappelleri JC, Cuervo J, Lang K, Mehta P, Mokgokong R, et al. Social preferences for health states associated with acute myeloid leukemia for patients undergoing treatment in the United Kingdom. *Health Qual Life Outcomes* 2018;16(1):66.

[60] Joshi N, Hensen M, Patel S, Xu W, Lasch K, Stolk E. Health state utilities for acute myeloid leukaemia: a time trade-off study. *PharmacoEcon* 2019;37(1):85-92.

[61] Leunis A, Redekop WK, Uyl-de Groot CA, Lowenberg B. Impaired health-related quality of life in acute myeloid leukemia survivors: a single-center study. *Eur J Haematol* 2014;93(3):198-206.

[62] Stein EM, Yang M, Guerin A, Gao W, Galebach P, Xiang CQ, et al. Assessing utility values for treatment-related health states of acute myeloid leukemia in the United States. *Health Qual Life Outcomes* 2018;16(1):193.

[63] Kurosawa S, Yamaguchi H, Yamaguchi T, Fukunaga K, Yui S, Wakita S, et al. Decision Analysis of Postremission Therapy in Cytogenetically Intermediate-Risk Acute Myeloid Leukemia: The Impact of FLT3 Internal Tandem Duplication, Nucleophosmin, and CCAAT/Enhancer Binding Protein Alpha. *Biol Blood Marrow Transplant* 2016;22(6):1125-1132.

[64] Szende A, Schaefer C, Goss TF, Heptinstall K, Knight R, Lubbert M, et al. Valuation of transfusion-free living in MDS: results of health utility interviews with patients. *Health Qual Life Outcomes* 2009;7:81.

[65] 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 Transplant* 2015;50(9):1241-9.

[66] Slovacek L, Slovackova B, Jebavy L, Macingova Z. Psychosocial, health and demographic characteristics of quality of life among patients with acute myeloid leukemia and malignant lymphoma who underwent autologous hematopoietic stem cell transplantation. *Sao Paulo Med J* 2007;125(6):359-61.

[67] Uyl-de Groot CA, Lowenberg B, Vellenga E, Suciu S, Willemze R, Rutten FF. Cost-effectiveness and quality-of-life assessment of GM-CSF as an adjunct to intensive remission induction chemotherapy in elderly patients with acute myeloid leukemia. *Br J Haematol* 1998;100(4):629-36.

[68] Stein E, Xie J, Duchesneau E, Bhattacharyya S, Vudumula U, Ndife B, et al. Cost effectiveness of midostaurin in the treatment of newly diagnosed FLT3-mutated acute myeloid leukemia in the United States. *PharmacoEcon* 2019;37(2):239-53.

[69] Pan F, Peng S, Fleurence R, Linnehan JE, Knopf K, Kim E. Economic analysis of decitabine versus best supportive care in the treatment of intermediate- and high-risk myelodysplastic syndromes from a US payer perspective. *Clin Ther* 2010;32(14):2444-56.

[70] Goss TF, Szende A, Schaefer C, Totten PJ, Knight R, Jadersten M, et al. Cost effectiveness of lenalidomide in the treatment of transfusion-dependent myelodysplastic syndromes in the United States. *Cancer Control* 2006;13 (Suppl):17-25.

[71] Grulke N, Albani C, Bailer H. Quality of life in patients before and after haematopoietic stem cell transplantation measured with the European Organization for Research and Treatment of Cancer (EORTC) Quality of Life Core Questionnaire QLQ-C30. *Bone Marrow Transplant* 2012;47(4):473-82.

[72] Perić Z, Desnica L, Duraković N, Ostojić A, Pulanić D, Serventi-Seiwerth R, et al. Which questionnaires should we use to evaluate quality of life in patients with chronic graft-vs-host disease? *Croat Med J* 2016;57(1):6-15.

[73] Lee S, Cook EF, Soiffer R, Antin JH. Development and validation of a scale to measure symptoms of chronic graft-versus-host disease. *Biol Blood Marrow Transplant* 2002;8(8):444-52.

[74] Lee SJ, Kim HT, Ho VT, Cutler C, Alyea EP, Soiffer RJ, et al. Quality of life associated with acute and chronic graft-versus-host disease. *Bone Marrow Transplant* 2006;38(4):305-10.

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

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

[77] Sutton L, Chevret S, Tournilhac O, Divine M, Leblond V, Corront B, et al. Autologous stem cell transplantation as a first-line treatment strategy for chronic lymphocytic leukemia: a multicenter, randomized, controlled trial from the SFGM-TC and GFLLC. *Blood* 2011;117(23):6109-19.

[78] Dakin H, Abel L, Burns R, Yang Y. Review and critical appraisal of studies mapping from quality of life or clinical measures to EQ-5D: an online database and application of the MAPS statement. *Health Qual Life Outcomes* 2018;16(1):31.

[79] Longworth L, Yang Y, Young T, Mulhern B, Hernandez Alava M, Mukuria C, et al. Use of generic and condition-specific measures of health-related quality of life in NICE decision-making: a systematic review, statistical modelling and survey. *Health Technol Assess* 2014;18(9):1-224.

[80] Khan I, Morris S. A non-linear beta-binomial regression model for mapping EORTC QLQ- C30 to the EQ-5D-3L in lung cancer patients: a comparison with existing approaches. *Health Qual Life Outcomes* 2014;12:163.

[81] Doble B, Lorgelly P. Mapping the EORTC QLQ-C30 onto the EQ-5D-3L: assessing the external validity of existing mapping algorithms. *Qual Life Res* 2016;25(4):891-911.

[82] Rowen D, Young T, Brazier J, Gaugris S. Comparison of generic, condition-specific, and mapped health state utility values for multiple myeloma cancer. *Value Health* 2012;15(8):1059-68.

[83] Hiramoto N, Kurosawa S, Tajima K, Okinaka K, Tada K, Kobayashi Y, et al. Positive impact of chronic graft-versus-host disease on the outcome of patients with de novo myelodysplastic syndrome after allogeneic hematopoietic cell transplantation: a single-center analysis of 115 patients. *Eur J Haematol* 2014;92(2):137-46.

[84] Longo M, Cohen D, Hood K, M R. Deriving an 'enhanced' EuroQoL from SF-36. Presented at the Health Economics Study Group (HESG) meeting, July 2000, Nottingham. 2000.

[85] Longworth L. *Estimating quality adjusted life years where health-related utility data are missing*: Brunel University; 2007.

[86] Rowen D, Brazier J, Roberts J. Mapping SF-36 onto the EQ-5D index: how reliable is the relationship? *Health Qual Life Outcomes* 2009;7:27.

[87] Franks P, Lubetkin EI, Gold MR, Tancredi DJ, Jia H. Mapping the SF-12 to the EuroQol EQ-5D Index in a national US sample. *Med Decis Making* 2004;24(3):247-54.

[88] Lawrence WF, Fleishman JA. Predicting EuroQoL EQ-5D preference scores from the SF-12 Health Survey in a nationally representative sample. *Med Decis Making* 2004;24(2):160-9.

[89] NHS Blood and Transplant. *NHS Blood and Transplant Annual Report and Accounts 2014/15 [Internet]*. Watford: NHS Blood and Transplant, 2015 [accessed 17.4.19]. 80p. Available from: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/443179/NHSBT_report_2014-15.pdf

[90] Medac. *Clinical phase III trial to compare treosulfan-based conditioning with busulfan-based reduced-intensity conditioning (RIC) prior to allogeneic haematopoietic stem cell transplantation in patients with AML or MDS considered ineligible to standard conditioning regimens. (Clinical trial report: treosulfan) Protocol MC-FludT.14/L - Part II [PDF supplied with the Company's submission]*: Medac, 15 Sep 2017. 1515p.

[91] Medicines Complete. British National Formulary [Internet]. London: The Royal Pharmaceutical Society, 2019 [accessed 1.4.19]. Available from: https://about.medicinescomplete.com/publication/british-national-formulary/

[92] Robin C, Hemery F, Dindorf C, Thillard J, Cabanne L, Redjoul R, et al. Economic burden of preemptive treatment of CMV infection after allogeneic stem cell transplantation: a retrospective study of 208 consecutive patients. *BMC Infect Dis* 2017;17(1):747.

[93] Esperou Hln, Brunot A, Roudot-Thoraval Fo, Buzyn A, Dhedin N, Rio B, et al. Predicting the Costs of Allogeneic Sibling Stem-Cell Transplantation: Results from a Prospective, Multicenter, French Study. *Transplantation* 2004;77(12):1854-1858.

[94] Khera N, Emmert A, Storer BE, Sandmaier BM, Alyea EP, Lee SJ. Costs of allogeneic hematopoietic cell transplantation using reduced intensity conditioning regimens. *Oncologist* 2014;19(6):639-44.

[95] Monitor, NHS England. *2016/17 National tariff payment system [Internet]*. London: Monitor, 2016 [accessed 1.4.19]. 91p. Available from:

https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/509697/2016-17_National_Tariff_Payment_System.pdf

[96] Medac. *Treosulfan with fludarabine for malignant disease before allogeneic stem cell transplant [ID1508]. Appendix M2 - OS & EFS results - AML [Excel document supplied with the company's submission]*: Medac, 2019

[97] Medac. *Treosulfan with fludarabine for malignant disease before allogeneic stem cell transplant [ID1508]. Appendix M3 - OS & EFS results - MDS [Excel document supplied with the company's submission]*: Medac, 2019

[98] Medac. *Treosulfan with fludarabine for malignant disease before allogeneic stem cell transplant [ID1508]. Appendix M4 - OS & EFS results - AML + MDS [Excel document supplied with the company's submission]*: Medac, 2019

[99] Troy JD, Atallah E, Geyer JT, Saber W. Myelodysplastic syndromes in the United States: an update for clinicians. *Ann Med* 2014;46(5):283-9.

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

[101] Medac. *Treosulfan with fludarabine for malignant disease before allogeneic stem cell transplant [ID1508]. Appendix M1 - Philips checklist validation [Excel document supplied with the company's submission]*: Medac, 2019

[102] Kaltenthaler E, Carroll C, Hill-McManus D, Scope A, Holmes M, Rice S, et al. The use of exploratory analyses within the National Institute for Health and Care Excellence single technology appraisal process: an evaluation and qualitative analysis. *Health Technol Assess* 2016;20(26):1-48.

Appendix 1: ERG goodness of fit assessment of parametric survival models: subgroup populations

The goodness of fit assessment was based on AIC/BIC values reported in Tables 5.4 to 5.7 and visual fit of the parametric curves vs. KM data as presented in the electronic model. The ERG assessment is summarised in tables below.

Based on Tables A1.1 to A1.4, the ERG concluded that, for the AML subgroup, the best candidate to model OS was the MCM lognormal distribution and to model EFS was the NMCM lognormal distribution. Note that it was assumed that the same type of distribution should be used for both treosulfan and busulfan. A gamma distribution could be used for EFS as an alternative.

Table A1.1: Overall goodness-of-fit assessment by the ERG: treosulfan EFS in the AML population

Parametric model	Visual fit with KM	AIC	BIC	Overall assessment [score]
Exponential	☹	☹	☹	☹ [1]
Weibull	☹	☹	☹	☹ [1]
Lognormal	☹	☹	☹	☹ [1]
Log-logistic	☹	☹	☹	☹ [1]
Gompertz	☺	☹	☹	😐 [1.7]
Gamma	😐	☺	☺	☺ [2.7]
MCM Weibull	😐	☹	☹	☹ [1.3]
MCM Lognormal	😐	☺	☺	☺ [2.7]
NMCM Weibull	😐	☹	☹	☹ [1.3]
NMCM Lognormal	😐	☺	☺	☺ [2.7]

Visual fit with KM: Based on ERG assessment.
AIC: Define $\Delta i = AICi - AICmin$, where AICi is the AIC of the i-th model, and AICmin is the lowest AIC, obtained for Gamma: ☺ = $\Delta i < 4$; 😐 = $4 < \Delta i < 7$; ☹ = $\Delta i > 10$.
BIC: Define $\Delta i = BICi - BICmin$, where BICi is the BIC of the i-th model, and BICmin is the lowest BIC, obtained for Gamma: ☺ = $\Delta i < 10$; ☹ = $\Delta i > 10$.
Overall score: average from ☺ = 3 points; 😐 = 2 points; ☹ = 1 point.
AIC = Akaike information criterion; BIC = Bayesian information criterion; ERG = Evidence Review Group; KM = Kaplan-Meier

Table A1.2: Overall goodness-of-fit assessment by the ERG: busulfan EFS in the AML population

Parametric model	Visual fit with KM	AIC	BIC	Overall assessment [score]
Exponential	☹	☹	☹	☹ [1]
Weibull	☹	☹	☹	☹ [1]
Lognormal	☹	☹	☹	☹ [1]
Log-logistic	☹	☹	☹	☹ [1]
Gompertz	☺	☹	☹	😐 [1.7]
Gamma	😐	☺	☺	☺ [2.7]
MCM Weibull	😐	☹	☹	☹ [1.3]

MCM Lognormal	😐	☺	☺	☺ [2.7]
NMCM Weibull	☺	☹	☹	😐 [1.7]
NMCM Lognormal	☺	☺	☺	☺ [3]

Visual fit with KM: Based on company's assessment. The ERG is more neutral regarding this assessment.
AIC: Define Δi = AICi − AICmin, where AICi is the AIC of the i-th model, and AICmin is the lowest AIC, obtained for NMCM lognormal: ☺ = Δi < 4; 😐 = 4 < Δi < 7; ☹ = Δi >10.
BIC: Define Δi = BICi − BICmin, where BICi is the BIC of the i-th model, and BICmin is the lowest BIC, obtained for NMCM lognormal: ☺ = Δi < 10; ☹ = Δi >10.
Overall score: average from ☺ = 3 points; 😐 = 2 points; ☹ = 1 point.
AIC = Akaike information criterion; BIC = Bayesian information criterion; ERG = Evidence Review Group; KM = Kaplan-Meier

Table A1.3: Overall goodness-of-fit assessment by the ERG: treosulfan OS in the AML population

Parametric model	Visual fit with KM	AIC	BIC	Overall assessment [score]
Exponential	☹	☹	☹	☹ [1]
Weibull	☹	😐	☹	☹ [1.3]
Lognormal	☹	☺	☺	😐 [2.3]
Log-logistic	☹	😐	☺	😐 [2]
Gompertz	☺	☺	☺	☺ [3]
Gamma	😐	☺	☺	☺ [2.7]
MCM Weibull	☺	😐	☺	☺ [2.7]
MCM Lognormal	☺	☺	☺	☺ [3]
NMCM Weibull	☺	😐	☺	☺ [2.7]
NMCM Lognormal	☺	☺	☺	☺ [3]

Visual fit with KM: Based on company's assessment. The ERG is more neutral regarding this assessment.
AIC: Define Δi = AICi − AICmin, where AICi is the AIC of the i-th model, and AICmin is the lowest AIC, obtained for Gamma: ☺ = Δi < 4; 😐 = 4 < Δi < 7; ☹ = Δi >10.
BIC: Define Δi = BICi − BICmin, where BICi is the BIC of the i-th model, and BICmin is the lowest BIC, obtained for Lognormal: ☺ = Δi < 10; ☹ = Δi >10.
Overall score: average from ☺ = 3 points; 😐 = 2 points; ☹ = 1 point.
AIC = Akaike information criterion; BIC = Bayesian information criterion; ERG = Evidence Review Group; KM = Kaplan-Meier

Table A1.4: Overall goodness-of-fit assessment by the ERG: busulfan OS in the AML population

Parametric model	Visual fit with KM	AIC	BIC	Overall assessment [score]
Exponential	☹	☹	☹	☹ [1]
Weibull	☹	☹	☹	☹ [1]
Lognormal	☹	☹	☹	☹ [1]
Log-logistic	☹	☹	☹	☹ [1]
Gompertz	😐	☹	☺	😐 [2]
Gamma	☹	😐	☺	😐 [2]
MCM Weibull	☺	☺	☺	☺ [3]

MCM Lognormal	☺	☺	☺	☺ [3]
NMCM Weibull	☺	☺	☺	☺ [3]
NMCM Lognormal	☺ ln better	☺	☺	☺ [3]

Visual fit with KM: Based on company's assessment. The ERG is more neutral regarding this assessment.
AIC: Define Δi = AICi − AICmin, where AICi is the AIC of the i-th model, and AICmin is the lowest AIC, obtained for MCM lognormal: ☺ = Δi < 4; 😐 = 4 < Δi < 7; ☹ = Δi >10.
BIC: Define Δi = BICi − BICmin, where BICi is the BIC of the i-th model, and BICmin is the lowest BIC, obtained for MCM lognormal: ☺ = Δi < 10; ☹ = Δi >10.
Overall score: average from ☺ = 3 points; 😐 = 2 points; ☹ = 1 point.
AIC = Akaike information criterion; BIC = Bayesian information criterion; ERG = Evidence Review Group; KM = Kaplan-Meier

Based on Tables A1.5 to A1.8, the ERG concluded that, for the MDS subgroup, the best candidate to model OS was the NMCM Weibull distribution and to model EFS was the MCM lognormal distribution. Note that it was assumed that the same type of distribution should be used for both treosulfan and busulfan. The curves chosen for EFS resulted in an overestimation for treosulfan. The only way to avoid this issue was assuming different distributions for treosulfan and busulfan. In that case, a lognormal distribution should be used for treosulfan and an MCM Weibull for busulfan. It should be noted though that the overall visual fit for treosulfan EFS and busulfan OS was poor suggesting that there is uncertainty associated with the parametric curves used to model EFS and OS.

Table A1.5: Overall goodness-of-fit assessment by the ERG: treosulfan EFS in the MDS population

Parametric model	Visual fit with KM	AIC	BIC	Overall assessment [score]
Exponential	☹	☹	☺	😐 [1.7]
Weibull	😐	😐	☺	😐 [2.3]
Lognormal	😐	☺	☺	☺ [2.7]
Log-logistic	😐	😐	☺	😐 [2.3]
Gompertz	😐	😐	☺	😐 [2.3]
Gamma	😐	☺	☺	☺ [2.7]
MCM Weibull	😐	😐	☺	😐 [2.3]
MCM Lognormal	😐	☺	☺	☺ [2.7]
NMCM Weibull	😐	😐	☺	😐 [2.3]
NMCM Lognormal	😐	☺	☺	☺ [2.7]

Visual fit with KM: Based on ERG assessment.
AIC: Define Δi = AICi − AICmin, where AICi is the AIC of the i-th model, and AICmin is the lowest AIC, obtained for Gamma: ☺ = Δi < 4; 😐 = 4 < Δi < 7; ☹ = Δi >10.
BIC: Define Δi = BICi − BICmin, where BICi is the BIC of the i-th model, and BICmin is the lowest BIC, obtained for Gamma: ☺ = Δi < 10; ☹ = Δi >10.
Overall score: average from ☺ = 3 points; 😐 = 2 points; ☹ = 1 point.
AIC = Akaike information criterion; BIC = Bayesian information criterion; ERG = Evidence Review Group; KM = Kaplan-Meier

Table A1.6: Overall goodness-of-fit assessment by the ERG: busulfan EFS in the MDS population

Parametric model	Visual fit with KM	AIC	BIC	Overall assessment [score]
Exponential	☹	☹	☹	☹ [1]
Weibull	☹	☹	☹	☹ [1]
Lognormal	☹	☹	☺	☹ [1.3]
Log-logistic	☹	☹	☹	☹ [1]
Gompertz	☺	☺	☺	☺ [3]
Gamma	☹	☺	☺	😐 [2.3]
MCM Weibull	☺	☺	☺	☺ [3]
MCM Lognormal	😐	☺	☺	☺ [2.7]
NMCM Weibull	☺	☺	☺	☺ [3]
NMCM Lognormal	😐	☺	☺	☺ [2.7]

Visual fit with KM: Based on company's assessment. The ERG is more neutral regarding this assessment.
AIC: Define $\Delta i = AICi - AICmin$, where AICi is the AIC of the i-th model, and AICmin is the lowest AIC, obtained for MCM lognormal: ☺ = $\Delta i < 4$; 😐 = $4 < \Delta i < 7$; ☹ = $\Delta i > 10$.
BIC: Define $\Delta i = BICi - BICmin$, where BICi is the BIC of the i-th model, and BICmin is the lowest BIC, obtained for MCM lognormal: ☺ = $\Delta i < 10$; ☹ = $\Delta i > 10$.
Overall score: average from ☺ = 3 points; 😐 = 2 points; ☹ = 1 point.
AIC = Akaike information criterion; BIC = Bayesian information criterion; ERG = Evidence Review Group; KM = Kaplan-Meier

Table A1.7: Overall goodness-of-fit assessment by the ERG: treosulfan OS in the MDS population

Parametric model	Visual fit with KM	AIC	BIC	Overall assessment [score]
Exponential	☹	😐	☺	😐 [2]
Weibull	☹	☺	☺	😐 [2.3]
Lognormal	☹	☺	☺	😐 [2.3]
Log-logistic	☹	☺	☺	😐 [2.3]
Gompertz	😐	☺	☺	☺ [2.7]
Gamma	☹	☺	☺	😐 [2.3]
MCM Weibull	☺	😐	☺	☺ [2.7]
MCM Lognormal	😐	☺	☺	☺ [2.7]
NMCM Weibull	☺	☺	☺	☺ [3]
NMCM Lognormal	😐	☺	☺	☺ [2.7]

Visual fit with KM: Based on company's assessment. The ERG is more neutral regarding this assessment.
AIC: Define $\Delta i = AICi - AICmin$, where AICi is the AIC of the i-th model, and AICmin is the lowest AIC, obtained for Gamma: ☺ = $\Delta i < 4$; 😐 = $4 < \Delta i < 7$; ☹ = $\Delta i > 10$.
BIC: Define $\Delta i = BICi - BICmin$, where BICi is the BIC of the i-th model, and BICmin is the lowest BIC, obtained for Lognormal: ☺ = $\Delta i < 10$; ☹ = $\Delta i > 10$.
Overall score: average from ☺ = 3 points; 😐 = 2 points; ☹ = 1 point.
AIC = Akaike information criterion; BIC = Bayesian information criterion; ERG = Evidence Review Group; KM = Kaplan-Meier

Table A1.8: Overall goodness-of-fit assessment by the ERG: busulfan OS in the MDS population

Parametric model	Visual fit with KM	AIC	BIC	Overall assessment [score]
Exponential	☹	☹	☹	☹ [1]
Weibull	☹	☹	☹	☹ [1]
Lognormal	😐	😐	☺	😐 [2.3]
Log-logistic	😐	☹	☺	😐 [2]
Gompertz	😐	☺	☺	☺ [2.7]
Gamma	😐	☺	☺	☺ [2.7]
MCM Weibull	😐	☺	☺	☺ [2.7]
MCM Lognormal	😐	☺	☺	☺ [2.7]
NMCM Weibull	😐	☺	☺	☺ [2.7]
NMCM Lognormal	😐	☺	☺	☺ [2.7]

Visual fit with KM: Based on company's assessment. The ERG is more neutral regarding this assessment.
AIC: Define $\Delta i = AICi - AICmin$, where AICi is the AIC of the i-th model, and AICmin is the lowest AIC, obtained for NMCM Weibull: ☺ = $\Delta i < 4$; 😐 = $4 < \Delta i < 7$; ☹ = $\Delta i > 10$.
BIC: Define $\Delta i = BICi - BICmin$, where BICi is the BIC of the i-th model, and BICmin is the lowest BIC, obtained for Gompertz: ☺ = $\Delta i < 10$; ☹ = $\Delta i > 10$.
Overall score: average from ☺ = 3 points; 😐 = 2 points; ☹ = 1 point.
AIC = Akaike information criterion; BIC = Bayesian information criterion; ERG = Evidence Review Group; KM = Kaplan-Meier

Appendix 2: Explanation of the changes made by the ERG to the implementation of OS and EFS in the electronic model

Correction 1:

The company used mortality rates instead of transition probabilities to calculate OS state membership. The method used by the company is incorrect. However, the differences with the corrected one are minor because the mortality rates (R) are small and the transition probabilities (calculated as 1-EXP(-R)) are similar to the mortality rates. A full explanation is included in the ERG revised version of the electronic model.

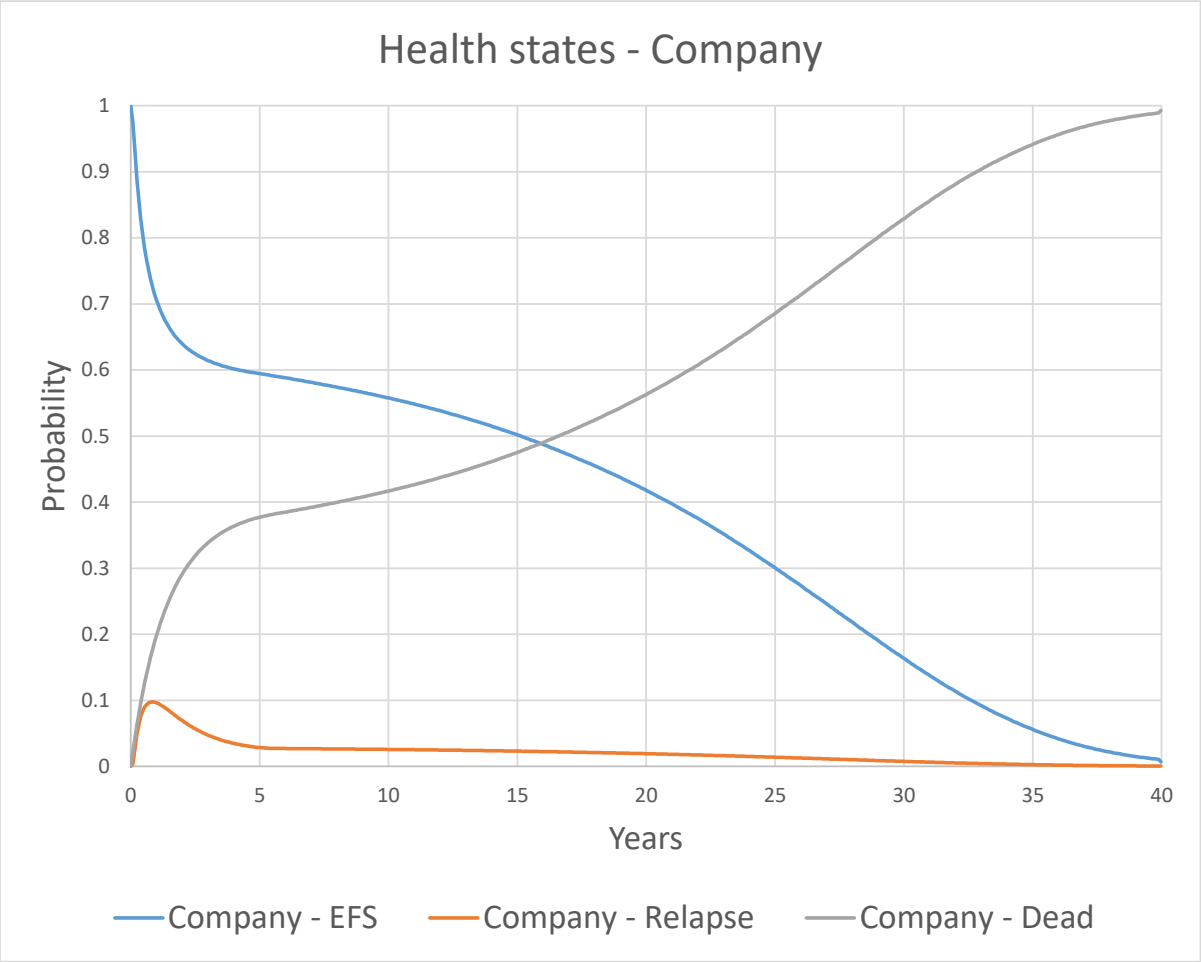
Correction 2:

The company applied mortality rates from HSCT-adjusted life tables to amend OS, when OS rates were the lowest. This seems appropriate to avoid patients dying at a lower rate than the HSCT-adjusted general population. However, the company applied the same approach to amend EFS. However, since EFS is defined by three events (relapse, graft failure and death), the ERG prefers using unadjusted EFS, allowing thus OS and EFS curves crossing. The only additional constraint is EFS < OS, so that at the time the curves cross, both OS and EFS are determined by the OS curve (implying that there is no relapse).

The health state membership for the treosulfan pooled population following the company and the ERG approach to EFS are shown in Figure A2.1 and A2.2, respectively. It can be seen that the main difference between the two approaches is in the relapse health state. The company approach allows relapses over the complete lifetime, while the ERG approach allows relapses for approximately 10 years (so also after the "cure point"). This duration depends on the curves selected to model OS and EFS.

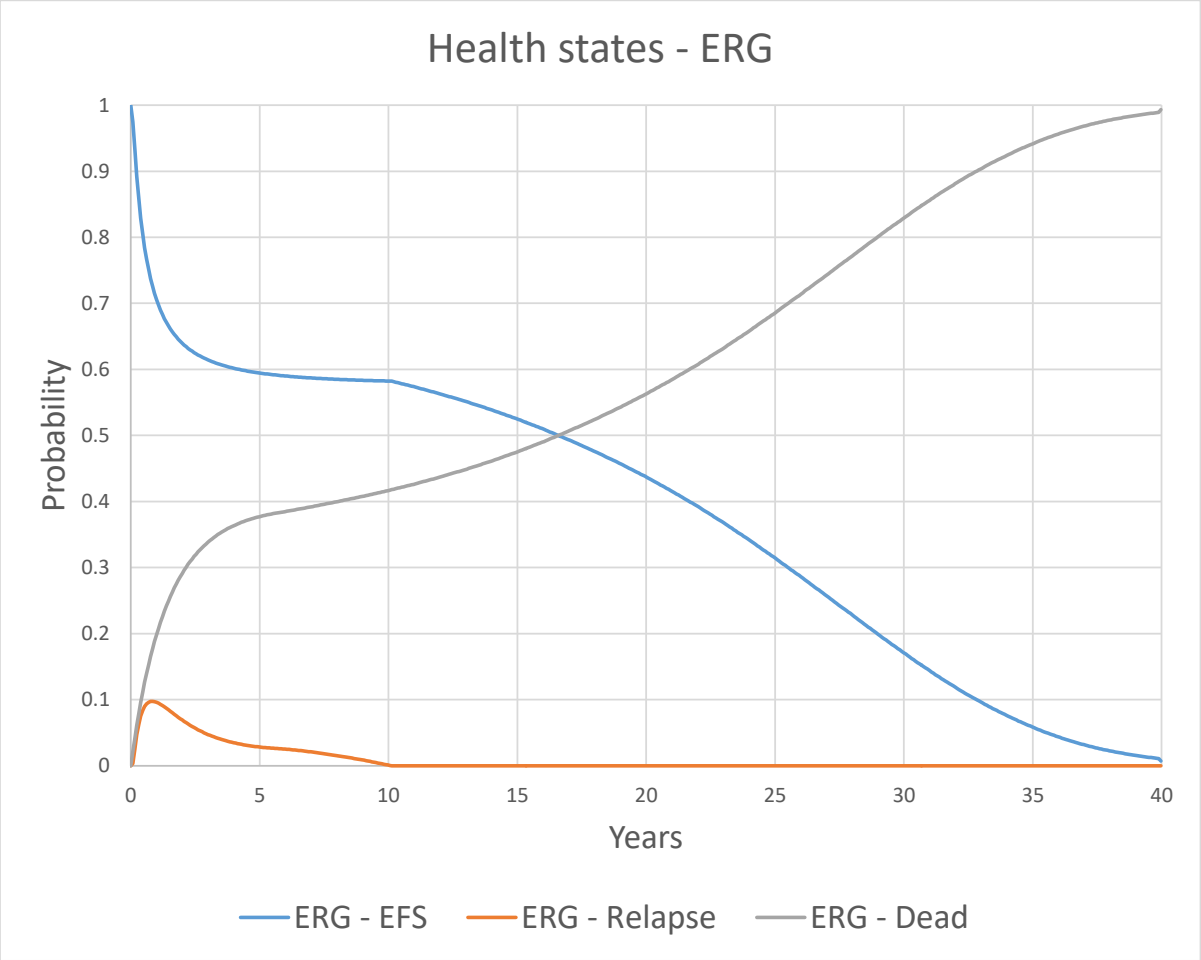
In response to the clarification question C2,[12] the company indicated that even though the curves used in the base-case analysis (NMCM lognormal) were plausible, one of the clinical experts raised an initial concern around "*the slight overestimation of patients in the relapse health state in the long-term*". Furthermore, as can be seen in Figure A2.3, extrapolations of NMCM lognormal distribution, as estimated by the company regression model (with no adjustments made), cross after approximately 7 years. Based on this, the ERG approach seems plausible and might overcome the issue raised by the clinical expert. A full explanation is included in the ERG revised version of the electronic model.

Figure A2.1: Health state membership (company approach to EFS): treosulfan pooled population



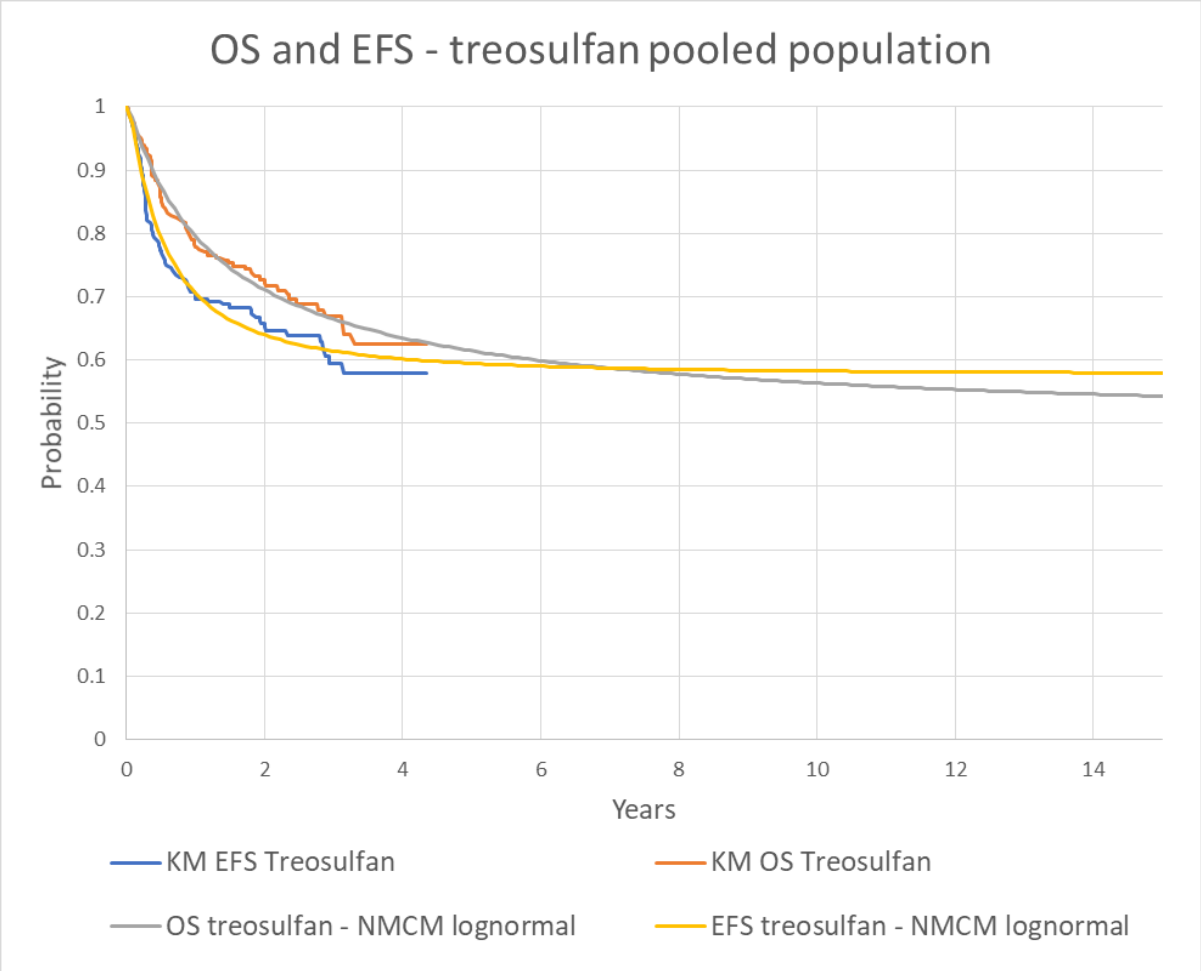
EFS = event-free survival

Figure A2.2: Health state membership (ERG approach to EFS): treosulfan pooled population



EFS = event-free survival; ERG = Evidence Review Group

Figure A2.3: KM curves and NMCM lognormal extrapolations of OS and EFS in the treosulfan pooled population



EFS = event-free survival; KM = Kaplan-Meier; NMCM = non-mixture-cure model; OS = overall survival

National Institute for Health and Care Excellence
Centre for Health Technology Evaluation

ERG report – factual accuracy check

Treosulfan with fludarabine for malignant disease before allogeneic stem cell transplant [ID1508

You are asked to check the ERG report to ensure there are no factual inaccuracies contained within it.

If you do identify any factual inaccuracies you must inform NICE by **5pm on Wednesday 7 August 2019** using the below comments table. All factual errors will be highlighted in a report and presented to the Appraisal Committee and will subsequently be published on the NICE website with the committee papers.

The factual accuracy check form should act as a method of detailing any inaccuracies found and how and why they should be corrected.

Issue 1 – Comparators

Description of problem	Description of proposed amendment	Justification for amendment	ERG comment
Section 3.3 - Comparators, page 30 In the ERG report, it states "In its definition of the decision problem (Table 3.1), the company indicates that the comparator(s) are "as final scope", however, the comparators listed in Table 3.1 do not include regimens with thiotepa and the reduced intensity conditioning regimen, melphalan plus fludarabine, is also not explicitly listed." It then goes on to discuss part of our response to the ERGs question on the comparators, however it excludes are rationale for not including thiotepa which was detailed in our response to question A1	Our response to question A1 should be provided: "Treosulfan and busulfan are stem cell depleting agents; however, thiotepa does not enhance stem cell depletion. The primary treatment effect of thiotepa within a conditioning regimen is to provide additional immunosuppression thereby preventing graft rejection.1 Therefore, thiotepa is always combined with a stem cell toxic treatment modality (e.g. treosulfan, busulfan, melphalan or total body irradiation). Therefore thiotepa was effectively included in the search, as an adjunct to the primary cell depleting agents included in the search terms. Today, thiotepa is usually combined with treosulfan or busulfan, especially in paediatric patients. In medac's trial MC-FludT.17/M, on investigators discretion, for additional immune-suppression, subjects could receive thiotepa in 2 single doses of 5 mg/kg given on Day -2. Indeed, 65 out of 70 subjects received such additional treatment. Furthermore, medac is currently running a trial comparing conditioning with treosulfan with fludarabine ± thiotepa versus busulfan with fludarabine ± thiotepa in paediatric patients with non-malignant diseases. In	The ERG report focusses only on our response to the reduced intensity conditioning regimen without mentioning our response to the issue of thiotepa. We think it also important to clarify the issue of why thiotepa was not utilised as a comparator in our submission.	Not a factual error. Question A1 concerned the lack of search terms for thiotepa in the literature searches used to inform the systematic literature review (SLR) submitted by the company. The ERG accepts the explanation, provided by the company and reproduced in this table, that thiotepa is used as an adjunctive treatment and hence any studies of regimens which included thiotepa would be captured by searches for the stem cell toxic treatment modalities to which it is added (e.g. treosulfan, busulfan, melphalan or total body irradiation). Therefore, section 4.1 of the ERG report did not include any criticism of the SLR on this basis. However, the issue of the definition of comparators is a separate one (addressed by clarification question B2). This question addresses the extent to which any comparator studies identified by the SLR were considered relevant for inclusion in the submission and whether the criteria for inclusion matched the

	this trial, the additional thiotepa can be administered as investigator's choice for a given patient (MC-FludT.16/NM) and again the majority of patients treated so far received additional thiotepa in both groups. Thiotepa cannot therefore be considered as an alternative or comparator to treosulfan but as an important adjunct to the treosulfan with fludarabine or busulfan with fludarabine regimen."		scope. The ERG report reproduced the company's response to question B2, with the exception of the following paragraph: "Because of the short timescale from referral to submission, without any prior warning, it was necessary to commission the work required for the submission on the basis of the post-referral Scope, and we were not aware that a revised Final Scope had been issued. Table 1 in document B should list the comparators as the final scope but with the removal of thiotepa (as discussed in our response to question A1)." As question A1 refers to the literature search (as detailed above) and our report did not include any criticism of the literature search on this point, we do not believe that the addition of the detailed response to question A1 would add any relevant information.

Issue 2 – Model structure description

Description of problem	Description of proposed amendment	Justification for amendment	ERG comment
Section 5.2.2 – Model structure,	Reference to the starting model health state	Description of model structure does not include the starting health sate	The ERG agrees with the company and the amendment

page 87 In the ERG report, it states that "The simulation assumes that all patients have received an HSCT. Therefore, all patients start in the remission health state. From there, patients can either relapse or die."	should be included: "The simulation assumes that all patients have received an HSCT. Therefore, all patients start in the induction/HSCT health state. From there, patients can enter post-HSCT recovery (remission), relapse/progress or die."	(Induction/HSCT) which is used to apply appropriate costs/utilities to patients undergoing HSCT in the first cycle.	has been incorporated as suggested.

Issue 3 – ERG scenario analysis description (utilities)

Description of problem	Description of proposed amendment	Justification for amendment	ERG comment
Section 5.2.8.4 – Health state utility scenarios, page 111-113 In the ERG report, utility-related scenario analyses conducted by the ERG are described in the section summarising the HRQoL data used in the cost-effectiveness analysis by the company.	Moving description of the ERG utility scenario analyses to the ERG comment section below in Section 5.2.8.4 or Section 7.1.3.6 (which summarises the utility scenarios explored by the ERG).	Scenarios explored by the ERG are not otherwise included in parts of the report where the ERG describes the company approach. Furthermore, utility scenario analyses explored by the company and ERG were similar, and as such moving this into sections which makes clear reference to ERG comments or ERG analyses provides clearer differentiation between scenario analyses explored by the company and the ERG.	The ERG agrees with the proposed amendment. Table 5.16 has been moved to the ERG comment of section 5.2.8.4 and mention of planned scenario analysis is removed from the 5.2.8.4 summary of the company approach

Issue 4 – Description of wastage and scenario analysis for busulfan dosing

Description of problem	Description of proposed amendment	Justification for amendment	ERG comment
Section 5.2.9.1 – ERG comment on treatment costs for treosulfan, page 117 In Section 5.2.9.1, the ERG states that "The high costs due to wastage of **treosulfan** are caused by a combination of 1) the assumption of 100% vial wastage (based on expert opinion), 2) the average weight of the trial patients (indicating a dosage that is just slightly over the available vial size), and 3) the only available vial size (identified by the BNF) being 60 mg. This results in nearly half of the total amount of **treosulfan** being wasted. The underlying rationale for the wastage was confirmed by the company in the response to the ERG's clarification questions. Following an updated summary of the product characteristics, a scenario (scenario 5 in the CS) was run where **treosulfan** dosage was based on a single daily dose. This scenario resulted in substantially reduced wastage (costs). However, this did not affect the overall conclusions of the	Replacing "**treosulfan**" with "**busulfan**" in the text: "The high costs due to wastage of **busulfan** are caused by a combination of 1) the assumption of 100% vial wastage (based on expert opinion), 2) the average weight of the trial patients (indicating a dosage that is just slightly over the available vial size), and 3) the only available vial size (identified by the BNF) being 60 mg. This results in nearly half of the total amount of **busulfan** being wasted. The underlying rationale for the wastage was confirmed by the company in the response to the ERG's clarification questions. Following an updated summary of the product characteristics, a scenario (scenario 5 in the CS) was run where **busulfan** dosage was based on a single daily dose. This scenario resulted in substantially reduced wastage (costs). However, this did not affect the overall conclusions of the economic analysis."	The ERG incorrectly describes wastage, wastage costs and scenario analysis attributable to busulfan as being applicable to treosulfan.	The ERG agrees with the company and the amendment has been incorporated as suggested.

economic analysis.”			

Issue 5 – Description of monitoring costs for long-term remission patients applied by the company

Description of problem	Description of proposed amendment	Justification for amendment	ERG comment
Section 5.2.9.2 – Event-free survival resource use costs, page 117 In Section 5.2.9.2, the ERG states that “Based on key opinion leader (KOL) input, after the initial two-year period the costs of HSCT patients who had not relapsed are those of three times monitoring, consisting of an outpatient haematology appointment, blood biochemistry, full blood counts and phlebotomy.”	Including more specific reference to the frequency of monitoring assumed in the company model: “Based on key opinion leader (KOL) input, after the initial two-year period the costs of HSCT patients who had not relapsed are those of three times monitoring **per year**, consisting of an outpatient haematology appointment, blood biochemistry, full blood counts and phlebotomy.”	Lack of clarity around the frequency of monitoring costs applied by the company.	The ERG agrees with the company and the amendment has been incorporated as suggested.

Issue 6 – Incorrect reference cited by the ERG for GvHD costs

Description of problem	Description of proposed amendment	Justification for amendment	ERG comment
Section 5.2.9.3 – Adverse event costs, pages 124-125 In Section 5.2.9.3 and Table 5.25, the ERG makes several references to a study by “Kloos et al 2017” as a source of GvHD costs identified in the literature	Replace all references to “Kloos et al 2017” in Section 5.2.9.3 with the correct reference (“Robin et al 2017”).	The ERG has cited the incorrect reference for a source of GvHD costs identified in the targeted literature review and used in the company submission model. Kloos et al 2017 does not appear to be cited elsewhere in the ERG	The ERG agrees with the company and the amendment has been incorporated as suggested.

search and utilitised in the company submission model basecase analysis. **Section 7.4 – Conclusions of the cost-effectiveness section, page 155** In Section 7.4, the ERG states that "Adverse event costs for extensive cGvHD were sourced from Kloos et al. 2017 and stage III/IV aGvHD costs." **Section 9 – References, page 166** In Section 9, the ERG lists the following citation (reference 92): "Kloos RQH, Uyl-de Groot CA, van Litsenburg RRL, Kaspers GJL, Pieters R, van der Sluis IM. A cost analysis of individualized asparaginase treatment in pediatric acute lymphoblastic leukemia. Pediatr Blood Cancer 2017;64(12)."	Change reference from "Kloos et al 2017" to "Robin et al 2017" in Section 7.4. Replace Kloos et al 2017 in the reference list with the citation for Robin et al 2017: Robin C, Hémery F, Dindorf C, Thillard J, Cabanne L, Redjoul R, Beckerich F, Rodriguez C, Pautas C, Toma A, Maury S, Durand-Zaleski I, Cordonnier C. Economic burden of preemptive treatment of CMV infection after allogeneic stem cell transplantation: a retrospective study of 208 consecutive patients. BMC Infect Dis. 2017 Dec 5;17(1):747.	report, and so can be directly replaced with Robin et al 2017.	

Issue 7 – Results of company deterministic sensitivity analysis

Description of problem	Description of proposed amendment	Justification for amendment	ERG comment
Section 6.2.2 – Deterministic sensitivity analysis, page 131	Change the last sentence of the paragraph to the following:	Lack of clarity around parameters that produced positive ICERs in the	The ERG agrees that the current wording may be

In Section 6.2.2, the ERG describes that "The company also presented the results of the one-way deterministic sensitivity analysis in terms of the ICER (not shown here). This confirmed the conclusions based on the INMB but also showed that for all parameters the ICERs were negative, indicating that treosulfan was dominating busulfan. The only exception was the meanlog coefficient of the busulfan OS distribution which in its upper limit resulted in an ICER of £20 per QALY, thus, below the £30,000 threshold."	"Aside from the meanlog coefficients of the busulfan EFS distribution and the treosulfan OS distribution, the only additional parameter that produced a positive ICER was the meanlog coefficient of the busulfan OS distribution, which in its **lower** limit resulted in an ICER of **£3,584** per QALY, thus, below the £30,000 threshold."	deterministic sensitivity analysis. Incorrect ICER value and limit described for the meanlog coefficient of the busulfan OS distribution.	confusing. The last sentence of the paragraph has been amended to read: "The only exceptions (with positive ICERs) were the meanlog coefficients of the treosulfan OS distribution, the busulfan EFS distribution and the busulfan OS distribution which resulted in upper bound ICERs of £34,821, £33,040 and £20 per QALY, respectively." However, the ERG does not agree that the incorrect value has been used. Figure 26 of the CS doc B shows the upper bound ICER for the busulfan OS meanlog coefficient to be £20, as does the original model submitted by the company. Therefore reference to this value is retained

Issue 8 – Assumption of 20% variation for the deterministic sensitivity analysis

Description of problem	Description of proposed amendment	Justification for amendment	ERG comment
Section 6.2.2 – Deterministic sensitivity analysis, page 132 In Section 6.2.2, the ERG describes that "The tornado diagram indicated that only the coefficients of the survival distributions might have some	Include reference to the company response to the question for clarification (C29), as well as approaches adopted by recent AML technology appraisal submissions and sections of the NICE reference case related to exploring uncertainty: "The tornado diagram indicated that only the coefficients of the survival distributions might	ERG comment regarding the deterministic sensitivity analysis does not include the company response to the clarification question on the choice of +/-20% variation. The ERG comment also does not	The ERG has added reference to the company's response to clarification regarding this issue as requested. However, the ERG still argue that a confidence interval approach represents better practice than

impact on decision uncertainty. While this might be the case, it should be noted that the one-way sensitivity analysis was conducted, allowing 20% variation from mean value for all parameters, which seems arbitrary and may not represent an equally plausible range of variation for all input parameters. Similar to the PSA described in the CS, the limits of (95%) confidence intervals should be used for each parameter to calculate the upper and lower bounds of the bars in the tornado diagram. Therefore, while the ERG agrees that the model results seem to be more sensitive to changes in the survival distributions, the ERG also considers that the tornado diagram presented above should be interpreted with caution."	have some impact on decision uncertainty. While this might be the case, it should be noted that the one-way sensitivity analysis was conducted, allowing 20% variation from mean value for all parameters, which seems arbitrary and may not represent an equally plausible range of variation for all input parameters. Similar to the PSA described in the CS, the limits of (95%) confidence intervals **could** be used for each parameter to calculate the upper and lower bounds of the bars in the tornado diagram. **However, the NICE reference case does not explicitly state recommendations for parameter variation for deterministic sensitivity analysis.** **In their response to clarification question C29, the company clarified that the choice of 20% variation for all parameters for the deterministic sensitivity analysis was chosen as the objective of the deterministic sensitivity analysis was to identify the most sensitive parameters in the model, and that they were concerned about producing misleading results from the DSA using 95% confidence intervals due to the covariance between survival function parameters. Furthermore, they cited that the choice of 20% variation for inputs for the deterministic sensitivity analysis was consistent with two recent AML appraisals (TA399, TA523).** **The choice of consistent parameter variation was also used in TA545, where variation of +/- 10% was applied to all parameters for the deterministic sensitivity analysis.**	comment on the choice of parameter variation adopted by the other recent AML appraisals identified by the company in the clarification response (TA399 and TA523), as well as the +/- 10% variation was applied for all inputs in another recent AML technology appraisal (TA545). In each of these respective technology appraisals, the ERG did not criticise the use of consistent parameter variation in the deterministic sensitivity analysis. Furthermore, the NICE reference case does not state explicitly what range of parameter variation should be adopted. It is important to note also that in Section 5.8.11 of the NICE reference case, the following text is stated, which appears broadly consistent with the concerns raised by the company regarding covariance between parameters: "The use of univariate and best- or worst-case sensitivity analysis is an important way of identifying parameters that may have a substantial impact on the cost-effectiveness results and of explaining the key drivers of the model. **However, such analyses become increasingly unhelpful in**	an arbitrary cut-off approach. Additionally, the ERG notes that just because this issue was not criticised in previous appraisals and is not explicitly mentioned in the NICE reference case does not mean that this reflects best practice. Therefore, the text has been amended as follows: "In their response to clarification question C29, the company clarified that the choice of 20% variation for all parameters for the deterministic sensitivity analysis was chosen as the objective of the deterministic sensitivity analysis was to identify the most sensitive parameters in the model, and that they were concerned about producing misleading results from the DSA using 95% confidence intervals due to the covariance between survival function parameters. Furthermore, they cited that the choice of 20% variation for inputs for the deterministic sensitivity analysis was consistent with two recent AML appraisals (TA399,

Section 7.4 – Conclusions of the cost effectiveness section, page 156 In section 7.4, the ERG provides similar commentary to that stated in Section 6.2.2: "The one-way sensitivity analysis was conducted, allowing 20% variation from mean value for all parameters. This seems arbitrary to the ERG and may not represent an equally plausible range of variation for all input parameters. Similar to the PSA described in the CS, the limits of (95%) confidence intervals should be used for each parameter to calculate the upper and lower bounds of the bars in the tornado diagram. Given the time constraints associated with this	**In each of these AML technology appraisals, the ERG for each appraisal did not comment on the choice of parameter variation adopted for the deterministic sensitivity analysis.** Therefore, while the ERG agrees that the model results seem to be more sensitive to changes in the survival distributions, **and that the approach to parameter variation adopted by the company is consistent with several recent AML technology appraisals (and does not appear inconsistent with the NICE reference case)**, the ERG also considers that the tornado diagram presented above should be interpreted with caution." Similar to Section 6.2.2, amend the text to reflect the company response, approaches adopted in recent AML appraisals, and what is stated in the NICE reference case, and adjust text that explicitly states the use of 95% confidence intervals as "correct": "The one-way sensitivity analysis was conducted, allowing 20% variation from mean value for all parameters. This seems arbitrary to the ERG and may not represent an equally plausible range of variation for all input parameters. Similar to the PSA described in the CS, the limits of (95%) confidence intervals **could** be used for each parameter to calculate the upper and lower bounds of the bars in the tornado diagram. **However, the ERG acknowledges that the choice of consistent parameter variation for**	**representing the combined effects of multiple sources of uncertainty as the number of parameters increase.** The use of probabilistic sensitivity analysis can allow a more comprehensive characterisation of the parameter uncertainty associated with all input parameters." As such, the company does not agree that use of 95% confidence intervals to determine the range of input parameter variation for deterministic sensitivity analysis is necessarily a more suitable approach, and that it is not necessarily aligned with the NICE reference case.	TA523)". While the ERG acknowledges the concern that covariance between survival function parameters may affect 95% confidence intervals, this argument only applies to a small number of parameters included in the DSA and the ERG still feel that the confidence interval approach represents better practice. This is because, while the +/- 20% approach undertaken by the company provides an idea of which parameters are having the biggest impact on model results, the results of the company tornado plot do not provide information on the impact of the actual uncertainty related to each parameter, which would be provided if an approach based on confidence intervals were used. Unfortunately, given the time constraints associated with this project, the ERG was not able to test the impact of using 95% confidence intervals for the deterministic sensitivity analysis. Therefore, while the ERG agrees that the model results seem to be more sensitive to changes in the

project, the ERG was not able correct this in the model."	**the deterministic sensitivity analysis is in line with several recent NICE technology appraisals for AML (TA399, TA523, TA545), and that covariance between input parameters, particularly when a large number of input variables are included, may affect the ability to appropriately interpret results of deterministic sensitivity analyses. Furthermore, the NICE reference case does not provide explicitly statements on what upper and lower limits should be adopted for deterministic sensitivity analysis.** Given the time constraints associated with this project, the ERG was not able **to test the impact of using 95% confidence intervals for the deterministic sensitivity analysis**."		survival distributions, the ERG also considers that the tornado diagram presented above should be interpreted with caution."

Issue 9 – ERG scenario analysis - plausibility of 1-year cure point

Description of problem	Description of proposed amendment	Justification for amendment	ERG comment
Section 7.2.2.3 – Additional scenario 3: changing duration of cure point, page 147 In Section 7.2.2.3, the ERG describes that "It can be seen that, only when the "cure point" was assumed to be one year, treosulfan did not dominate busulfan. In that scenario, the resulting ICER was £47,910 in the SW quadrant of the CE-plane;	Include an additional sentence that references the KOL feedback collected by the company: "It can be seen that, only when the "cure point" was assumed to be one year, treosulfan did not dominate busulfan. In that scenario, the resulting ICER was £47,910 in the SW quadrant of the CE-plane; thus, treosulfan produced less QALYs than busulfan but also at lower costs. **However, given the base case assumption of five years adopted by the company based on KOL feedback and used in the ERG base**	ERG summary of cure point scenario analysis does not include commentary on the plausibility of the results obtained. Given that a 5-year cure point was assumed by the company (based on KOL feedback) and adopted for the ERG base case analysis, the choice of a 1-year cure point does not appear to be a plausible assumption.	Not a factual error. The model provided the option to select cure time points between 1-5 years. The company submission did not discuss the plausibility of different cure time point assumptions and in order to assess plausibility the ERG would require additional evidence from the company,

thus, treosulfan produced less QALYs than busulfan but also at lower costs.”	**case analysis, a one-year cure point is likely to be an implausible assumption**.”	Discussion around plausibility of the scenarios explored would also be consistent with descriptions of other scenario analysis results (such additional scenario 5, Section 7.2.2.5), where the ERG comments on the plausibility of the scenarios explored.	preferably based on data other than expert opinion.

Issue 10 – Description of utilities used in the company base case model

Description of problem	Description of proposed amendment	Justification for amendment	ERG comment
Section 7.4 – Description of utilities used in the company base case model, page 154 In Section 7.4, the ERG provides the following summary of health state utility values used in the company base case analysis: “The company identified a range of sources of utility evidence; however, none met the NICE reference case preference for UK EQ-5D values. Therefore, in the base-case the company utilised health state utility values from Grulke et al. 2012, mapping from the QLQ-C30 to the UK EQ-5D	Include an additional sentence that describes how relapse/progression utilities were estimated: ““The company identified a range of sources of utility evidence; however, none met the NICE reference case preference for UK EQ-5D values. Therefore, in the base-case the company utilised health state utility values from Grulke et al. 2012, mapping from the QLQ-C30 to the UK EQ-5D utility values using an established and appropriate mapping algorithm developed by Proskorovsky et al. 2014. **Health state utilities for relapse/progression patients were estimated using a disutility multiplier estimated from TA399 based on mappings of QLQ-C30 data to EQ-5D using Proskorovsky et al. 2014, which was applied to the short-term post-HSCT recovery (≤6 months)**	The summary of utilities used in the company base case analysis does not include a description of how relapse utilities were estimated.	The ERG agrees with the company and the suggested amendment has been incorporated

utility values using an established and appropriate mapping algorithm developed by Proskorovsky et al. 2014. Base-case disutilities for GvHD were identified from Kurosawa et al. 2016 based on non-UK EQ-5D values. Disutilities for Grade 3 and 4 AEs were identified from TA399 based on mappings from QLQ-C30 data, which were mapped to UK EQ-5D values, again using the Proskorovsky et al. 2014 mapping algorithm. Utilities applied in the model were adjusted for age and gender using the methods of Ara and Brazier 2010."	**estimated from Grulke et al. 2012.** Base-case disutilities for GvHD were identified from Kurosawa et al. 201663 based on non-UK EQ-5D values. Disutilities for Grade 3 and 4 AEs were identified from TA39943 based on mappings from QLQ-C30 data, which were mapped to UK EQ-5D values, again using the Proskorovsky et al. 201475 mapping algorithm. Utilities applied in the model were adjusted for age and gender using the methods of Ara and Brazier 2010."		

Issue 11 – Description of medical resource use costs for post-HSCT recovery patients

Description of problem	Description of proposed amendment	Justification for amendment	ERG comment
Section 7.4 – Description of post-HSCT recovery costs used in the company base case model, page 155 In Section 7.4, the ERG summarises that: "Post-HSCT recovery costs in EFS for the **first 18 months** were	Correct the description as follows: "Post-HSCT recovery costs in EFS for the **first 24 months** were based on an NHS Blood and Transplant Analysis 2015. **After 24 months**, these costs were based on clinical expert opinion."	Application of post-HSCT recovery costs are described incorrectly.	The ERG agrees with the company and the amendment has been incorporated as suggested.

based on an NHS Blood and Transplant Analysis 2015. **Between months 18 and 24**, these costs were based on clinical expert opinion."			

Issue 12 – Spelling/grammatical errors

Description of problem	Description of proposed amendment	Justification for amendment	ERG comment
Section 5.2.9.3 – Adverse event costs, page 125 In Section 5.2.9.3, the ERG describes that "For costing investigations such as increased GGT and febrile neutropenia HRG codes used in **TA5231** were not available since the 2014/2015 version of the NHS reference costs."	Correct "TA5231" to "TA523".	Incorrect number cited for technology appraisal guidance.	The ERG agrees with the company and the amendment has been incorporated as suggested.

NICE National Institute for Health and Care Excellence

Technical engagement response form

Treosulfan with fludarabine for malignant disease before allogeneic stem cell transplant [ID1508]

As a stakeholder you have been invited to comment on the technical report for this appraisal. The technical report and stakeholders' responses are used by the appraisal committee to help it make decisions at the appraisal committee meeting. Usually, only unresolved or uncertain key issues will be discussed at the meeting.

We need your comments and feedback on the questions below. You do not have to answer every question. The text boxes will expand as you type. Please read the notes about completing this form. We cannot accept forms that are not filled in correctly. Your comments will be summarised and used by the technical team to amend or update the scientific judgement and rationale in the technical report.

Deadline for comments: **5pm, Thursday 10 October 2019**

Thank you for your time.

Please log in to your NICE Docs account to upload your completed form, as a Word document (not a PDF).

Notes on completing this form

- Please see the technical report which summarises the background and submitted evidence. This will provide context and describe the questions below in greater detail.
- Please do not embed documents (such as PDFs or tables) because this may lead to the information being mislaid or make the response unreadable. Please type information directly into the form.
- Do not include medical information about yourself or another person that could identify you or the other person.
- Do not use abbreviations.
- Do not include attachments such as journal articles, letters or leaflets. For copyright reasons, we will have to return forms that have attachments without reading them. You can resubmit your form without attachments, but it must be sent by the deadline.
- If you provide journal articles to support your comments, you must have copyright clearance for these articles.
- Combine all comments from your organisation (if applicable) into 1 response. We cannot accept more than 1 set of comments from each organisation.
- Please underline all confidential information, and separately highlight information that is submitted under 'commercial in confidence' in turquoise, all information submitted under 'academic in confidence' in yellow. If confidential information is submitted, please also send a second version of

your comments with that information replaced with the following text: 'academic/commercial in confidence information removed'. See the Guide to the processes of technology appraisal (sections 3.1.23 to 3.1.29) for more information.

We reserve the right to summarise and edit comments received during engagement, or not to publish them at all, if we consider the comments are too long, or publication would be unlawful or otherwise inappropriate.

Comments received during engagement are published in the interests of openness and transparency, and to promote understanding of how recommendations are developed. The comments are published as a record of the comments we received, and are not endorsed by NICE, its officers or advisory committees.

About you

Your name	[redacted]
Organisation name – stakeholder or respondent (if you are responding as an individual rather than a registered stakeholder please leave blank)	**medac Pharma UK (medac)**
Disclosure Please disclose any past or current, direct or indirect links to, or funding from, the tobacco industry.	**Nothing to disclose**

Questions for engagement

Issue 1: How does treosulfan differ from other alkylating agents sued in conditioning regimens? What is the main benefit of treosulfan?	
1. What clinical advantages does treosulfan have over other alkylating agents used in conditioning regimens for alloHSCT?	The main advantage of treosulfan (TREO)-based conditioning is its lower toxicity compared to other conditioning regimens; especially the busulfan/fludarabine (BU/FLU) regimen which is one of the most frequently used conditioning regimens in many countries. The improved tolerability results in a significant and clinically relevant reduced non-relapse/transplant-related mortality and increased event-free survival (+14.5%) as well as overall survival (+ 12.5%). Due to its excellent tolerability, TREO/FLU is the first myeloablative conditioning (MAC) regimen which can also be used in older and/or comorbid patients. It combines the advantages of the low toxicity of reduced intensity conditioning (RIC) regimens with the better disease control of MAC regimens.
2. What are the causes of death in the group of people who do not relapse (non-relapse mortality)- which in the company's submission is said to drive the OS benefit- are they non-cancer related deaths?	The main causes of non-relapse death include infections and graft-versus-host disease (GvHD). Fewer deaths from infections (TREO vs. BU: 9.3% vs. 14.1%) or GvHD (4.8% vs. 7.4%) were seen with TREO/FLU compared to BU/FLU conditioning (both causes combined: 13.9% vs. 21.5%). The cumulative incidence of severe (grade III/IV) acute GvHD (6.4% vs. 8.1%) and especially extensive chronic GvHD (19.8% vs. 28.6% at 2 years; HR 0.71 [95% CI 0.48-1.04]) was lower in the TREO-treated patients.
3. What is the difference between transplantation-related mortality (TRM) and non-relapse mortality (NRM)?	Non-relapse mortality was defined as the probability of dying without previous occurrence of a relapse or progression of the underlying malignant disease. TRM was defined as all deaths occurring due to one of the following main causes: GvHD, cardiac toxicity, pulmonary toxicity, interstitial pneumonitis, haemorrhage, hepatic sinusoidal obstruction syndrome (HSOS), skin and subcutaneous tissue disorders, Epstein-Barr virus proliferative disease, renal failure, gastro-intestinal toxicity, rejection/poor graft function, central nervous system toxicity, multiple organ failure, infections, and other haematopoietic stem cell transplantation (HSCT)-related causes.

	In contrast to NRM, TRM also includes those patients who died from treatment while having a relapse/progression. Therefore, TRM is slightly higher than NRM.
4. The main driver of the model is delay to relapse (rather than non-relapse death which gives the OS benefit), but the relapse rates appear similar between treosulfan and busulfan, is there any time to relapse data to clarify this point?	Differences between treosulfan and busulfan in the economic model are driven by differences between the overall survival (OS) and event-free survival (EFS) curves, which determine the proportion of patients in the relapse/progression/graft failure health state over time. Events were defined as relapse of disease/progression, graft failure or death (whichever occurred first). EFS was significantly improved for treosulfan in the MC-FludT.14/L Trial II (Hazard Ratio (HR) [95% Confidence Interval (CI)] = 0.64 [0.49, 0.84]). The EFS difference was driven by fewer death events (13.1% vs. 19.8%), fewer relapse/progression events (22.8% vs. 25.4%), as well as fewer graft failure events (0.4% vs. 3.2%) in the TREO group.[1] Furthermore, mean time to relapse/progression (235.6 days vs. 212.5 days) and mean time to relapse/progression or graft failure (232.3 days vs. 203.4 days) were longer in the TREO group.[1] While delay in time to relapse/progression or graft failure contributed to differences in relapse/progression or graft failure health state membership in the model, time spent in the relapse/progression or graft failure health state is also a key driver in the model (see response to Issue 1, Question 5).
5. The modelling shows that treosulfan approximately halved the time in the relapse/progression health state compared to busulfan. How does this relate to the evidence from the trial?	As stated above in response to Issue 1, Question 4, differences in relapse/progression incidence as well as graft failures were observed in the MC-FludT.14/L Trial II. This generated larger differences in the OS and EFS curves for busulfan which define membership of the relapse/progression health state in the economic model. OS and EFS are also extrapolated beyond the duration of the clinical trial using parametric survival models and a functional cure assumption at 5 years. As the difference between the OS and EFS is larger at the end of the trial for busulfan compared to treosulfan, differences in OS and EFS observed in the clinical trial are then preserved beyond the duration of the clinical trial and for a longer period of time for busulfan compared to treosulfan. In the Evidence Review Group's (ERG's) preferred version of the economic analysis, the adjusted life table estimates are applied to the OS curves only. In this case, OS becomes equal to EFS at

	approximately 8 years for treosulfan compared to 13.3 years for busulfan.
Issue 2: To what extent can the results of the trial be extrapolated to a broader population of patients with other malignancies?	
1. Is it plausible to assume that the data from MC-FludT.14/L Trial II is generalisable to a broader population of people with malignant disease requiring conditioning treatment, specifically people with malignant diseases other than acute myeloid leukaemia (AML) and myelodysplastic syndromes (MDS)?	Yes, the data from MC-FludT.14/L Trial II is generalisable to a broader population of people with malignant disease requiring conditioning therapy. Treosulfan's conditioning effect is we believe disease-agnostic. Evidence has been provided within the large retrospective European Society for Blood and Marrow Transplantation (EBMT) registry study of patients with multiple myeloma.[2] Furthermore, Gran et al found that upfront treatment with TREO/FLU showed superior OS and NRM at 5 years compared to other RIC or MAC regimens.[2] The efficacy of TREO-based conditioning has also been shown in other malignancies such as lymphoid malignancies, multiple myeloma or chronic myeloid leukaemia (see data within the Haematologists' position paper submitted as Appendix L.7), which was written by six clinical transplant experts). As discussed above, the improved efficacy of TREO is mainly due to the reduced toxicity of this regimen compared to alternative MAC regimens. The toxicity of a conditioning regimen does not depend on the underlying malignancy.
2. To what extent could the treatment-effect on endpoints such as time to relapse and death observed in the trial be relevant to the broader population of people with malignant disease (the Kaplan-Meier estimates of event-free survival [EFS] are reported in Appendix 3)?	For the determination of EFS, events were defined as relapse of disease, graft failure or death (whichever occurred first). Since graft failures in patients with malignancies are rare (more frequent in non-malignant diseases), the main drivers of EFS are TRM and relapse of the underlying malignancy. The toxicity of the conditioning regimen is independent from the underlying malignancy, and is determined more by the comorbidity status and age of the patients. In contrast, the relapse rate depends more on the disease status at transplant, i.e. whether the patient is in first or later complete remission (CR1/> CR1), or has residual disease, or is heavily pre-treated; however, using a MAC regimen instead of RIC may reduce the risk for relapse. It should be remembered that TREO/FLU is a MAC regimen, although it may also be referred to as a reduced-toxicity conditioning regimen (RTC).
3. In clinical practice, is the efficacy of conditioning treatments such as treosulfan expected to be the	Yes, we believe the efficacy of treosulfan will be the same irrespective of the underlying malignant

same irrespective of the underlying malignant disease?	disease. The main purpose of conditioning is to eliminate the self-renewing capacity of the patient's own haematopoiesis, and to suppress the recipient's immune system in order to allow engraftment of donor stem cells. This is independent of the underlying disease (malignant or non-malignant).
4. In clinical practice, is the efficacy of conditioning treatment expected to be the same in patients for whom myeloablative conditioning (MAC) is not suitable and in patients who could have MAC?	Non-TREO-containing MAC regimens are not suited for older and/or comorbid patients because of their higher toxicity in such patients compared to younger patients without comorbidities. TREO/FLU is the first MAC regimen that can also be used in older and/or comorbid patients. TREO/FLU has also been studied in many trials (company-sponsored as well as investigator-initiated trials) that also included patients without risk factors for standard MAC regimens, which confirmed the low toxicity and efficacy of this regimen.[2–11]
5. In clinical practice, is the efficacy of conditioning treatment expected to be the same in adults as in children?	Outcomes in children with malignant diseases undergoing alloHSCT is usually better than in adults, most probably because children suffer less frequently from significant comorbidities. The efficacy of TREO-based conditioning was demonstrated in trial MC-FludT.17/M which included 70 children with acute myeloid leukaemia (AML), acute lymphoblastic leukaemia (ALL), myelodysplastic syndrome (MDS), or juvenile myelomonocytic leukaemia (JMML). A recently performed survival update of this trial showed a very low TRM (7.9% at 3 years) and a low cumulative incidence of relapse/progression (22.4%), resulting in an excellent overall survival (86.7% at 3 years) which is 20% better than with TREO/FLU in the pivotal trial MC-FludT.14/L Trial II in older/comorbid adults (66.8% at 3 years). (Please see Appendix 4 for details)
6. Is the MC-FludT.17/M trial sufficient to support the extension of the use of treosulfan to children?	Yes, the MC-FludT.14/L Trial II is sufficient to support the extension of use of treosulfan to children. The efficacy and safety of TREO-based conditioning in paediatric patients has been confirmed in many trials.[12–16] Data of TREO-based conditioning are available from the EBMT registry for a total of 1,416 patients which confirm its efficacy and safety for malignant as well as non-malignant transplant indications.[17,18] Current European guidelines already recommend TREO-based conditioning as an option for children with MDS,[19] as well as children with primary immunodeficiencies.[20]

7. Would additional cost-effectiveness analysis conducted with data from the MC-FludT.17/M (ongoing study in children) provide sufficient evidence to support the appraisal of treosulfan in children?	Study MC-FludT.17/M has now been completed and patients are currently in the follow-up period for survival status. The trial was a single-arm study and therefore only indirect comparisons could be made with other conditioning regimens. Survival results are better than in the pivotal study in adults as outlined above (question 5). Since for children the same BU/FLU regimen as in adults is frequently used, and the available evidence is less robust than that for adults, we decided not to perform an additional cost-effectiveness analysis.
Issue 3: The submission focuses on a low-intensity regimen	
1. Is the company's choice of comparator appropriate?	Our choice of comparator is appropriate because the BU/FLU regimen FB2 (RIC) is the most frequently used conditioning regimen in older and/or comorbid patients.
2. Is reduced intensity conditioning (RIC) the area with the largest unmet need?	Reduced intensity conditioning is not the area of largest unmet need. The area with the largest unmet need is the reduction of toxicity of myeloablative conditioning (MAC) regimens, i.e. to develop reduced-toxicity standard intensity (= myeloablative) conditioning regimens which can be used also in older and/or comorbid patients. Most adult patients with haematological malignancies are older than 60 years and often suffer from comorbidities (cardiac, liver, renal etc.). AlloHSCT is generally the only chance of cure for these patients; however, RIC regimens bear a risk of higher relapse rates.[21] TREO/FLU is a reduced-toxicity MAC regimen designed specifically to address the unmet need for reduced conditioning toxicity without impacting efficacy. Patients undergoing alloHSCT should not die from the toxicity of the conditioning regimen.
3. What are the conditioning treatments used in clinical practice in people at high risk for standard conditioning therapy (that is, not eligible to standard high-intensity MAC)?	There is no consensus over conditioning regimens for patients not eligible for standard high-intensity MAC regimens. The RIC regimens currently used most frequently across the world in patients not eligible for standard conditioning therapy include low-dose busulfan plus fludarabine (FB2) and melphalan plus fludarabine (FLU/MEL). TREO-based conditioning is increasingly being recognised by transplant experts as the future standard conditioning for such patients.
4. Are the comparator regimens all similarly effective?	Yes, results of an EBMT survey in 394 AML patients (median age 56 years; range, 21-76 years). suggest that although FLU/MEL provides better AML control than BU/FLU as an RIC regimen for

	alloHSCT, the two regimens provide similar survival.[22] However, multicentre randomised studies are needed to confirm these findings. Furthermore, results from the Center for International Blood and Marrow Transplant Research (CIBMTR) analysis with various non-TREO regimens (FLU + low or high dose BU, busulfan/cyclophosphamide (BU/CY), FLU/MEL) show nearly equivalent EFS and OS rates.[23]
5. For people at high risk for standard conditioning therapy who are eligible for reduced intensity regimens, is low dose busulfan with fludarabine as effective as melphalan with fludarabine? Is there any quantitative evidence (such as, clinical trials, registries) to support it? If not, can the clinical experts provide an estimate for how different it is?	BU/FLU is less toxic than FLU/MEL but less effective with respect to disease control. However, the net effect is a comparable overall survival rate (see also response to question 4). The submitted EBMT analysis supports this. 2-year overall survival rate with FLU/MEL in the matched (with TREO arm of the pivotal medac MC-FludT.14/L Trial II) cohort was 56.5% (95% CI 33.9-79.1) which is quite comparable to the results of the BU/FLU regimen in the trial (60.2%; 95% CI 54.0, 65.8).[24]
6. For people at lower risk who are eligible for standard regimens, what is the difference in time to event between standard high-intensity regimens and low dose busulfan? Is there any quantitative evidence available? If not, can the clinical experts provide an estimate for how different it is?	People at lower risk (age < 50 years, no significant comorbidities) are currently preferably treated with standard intensity (MAC) regimens because RIC regimens bear a higher risk of relapse. Currently, there is insufficient evidence to generalize about the comparative benefit of MAC versus RIC transplants for the major diseases (AML, MDS, ALL, NHL) in younger patients (< 50 years of age) without comorbidities. No prospective randomised study in a sufficient patient number exists and such a trial is unlikely to be performed as such patients tolerate MAC regimens and clinicians are unlikely to want to increase the relapse risk in these patients.
7. Do the regimens all have similar costs (e.g. acquisition cost, resource use for administration etc.)?	Estimated acquisition costs and total number of IV drug/TBI administrations for all regimens included in the final scope, based on the lowest per mg vial/tablet and excluding wastage are summarised below: **Conditioning regimen** | **Dosing** | **Total cost (excluding wastage)**

	Treosulfan + Fludarabine (MAC)	Treosulfan: 10 g/m^2 once daily over 3 days prior to alloHSCT Fludarabine: 30 mg/m^2 once daily over 5 days prior to alloHSCT	£2,569.44
	Busulfan + Fludarabine (RIC) - including Phenytoin	Busulfan: 4 x 0.8 mg/kg OR 1 x 3.2 mg/kg daily over 2 days prior to alloHSCT Fludarabine: 30 mg/m^2 once daily over 5 days prior to alloHSCT Phenytoin: 3 x 200 mg over 1 day then 3 x 100 mg over 3 days prior to alloHSCT	£1,796.33
	Busulfan + Fludarabine (RIC) - including Phenytoin	Busulfan: 4 x 0.8 mg/kg OR 1 x 3.2 mg/kg daily over 2 days prior to alloHSCT Fludarabine: 30 mg/m^2 once daily over 5 days prior to alloHSCT	£1,794.70
	Melphalan + Fludarabine (RIC) - including Alemtuzumab	Melphalan: Single dose of 140 mg/m^2 prior to alloHSCT Fludarabine: 30 mg/m^2 once daily over 5 days prior to alloHSCT Alemtuzumab: 20 mg/m^2 once daily over 5 days prior to alloHSCT	£2,601.90
	Melphalan + Fludarabine (RIC) - excluding Alemtuzumab	Melphalan: Single dose of 140 mg/m^2 prior to alloHSCT Fludarabine: 30 mg/m^2 once daily over 5 days prior to alloHSCT	£901.89
	Cyclophosphamide + TBI (MAC)	Cyclophosphamide: 60 mg/kg once daily over 5 days prior to alloHSCT TBI: 2 Gy fractions daily over 6 days (12 Gy total) prior to alloHSCT	£2,781.19
	Cyclophosphamide + Busulfan (MAC)	Cyclophosphamide: 60 mg/kg once daily over 5 days prior to alloHSCT Busulfan: 4 x 0.8 mg/kg daily OR 1 x 3.2 mg/kg daily over 4 days prior to alloHSCT	£3,373.68
	Cyclophosphamide + Thiotepa (MAC)	Cyclophosphamide: 60 mg/kg once daily over 5 days prior to alloHSCT Thiotepa: 5 mg/kg daily over 3 days prior to alloHSCT	£8,956.67
	Busulfan + Fludarabine (MAC) - including Thiotepa	Busulfan: 3.2 mg/kg once daily over 3 days prior to alloHSCT Fludarabine: 50 mg/m^2 once daily over 3 days prior to alloHSCT Thiotepa: 5 mg/kg once daily over 2 days	£8,515.19
	Busulfan + Fludarabine (MAC) - excluding Thiotepa	Busulfan: 3.2 mg/kg once daily over 3 days prior to alloHSCT Fludarabine: 50 mg/m^2 once daily over 3 days prior to alloHSCT	£2,612.47

	When excluding vial wastage, the costs of the different conditioning regimens (excluding other conditioning such as anti-thymocyte globulin (ATG), GvHD prophylaxis and the costs of cell harvesting and alloHSCT) range from £901.89 to £8,956.67. Alemtuzumab is often prescribed with melphalan and fludarabine regimens (especially in the UK), and as such, most patients on melphalan-based conditioning regimens would likely incur a treatment cost similar to treosulfan + fludarabine. Treosulfan + fludarabine is similar in cost to the least expensive MAC regimens (cyclophosphamide + TBI, busulfan + fludarabine without thiotepa). Individual conditioning regimens are unlikely to be significant drivers of overall costs given the short treatment period. In the economic analysis, costs of the specific differential regimens included (treosulfan + fludarabine, busulfan + fludarabine) constituted <3% of the overall costs, even when assuming 100% vial wastage. Clinical outcomes (such as OS and EFS) are likely to be more important drivers of economic differences between conditioning regimens. Regarding resource use for administration, the total number of treatment administrations varies from 6-11 for all treatments in the above table. This table suggests that differences in resource use requirements for administration of conditioning regimens may be relatively small compared to differences in cost generated through differences in clinical efficacy between conditioning regimens and are unlikely to represent a large proportion of the overall costs.
Issue 4: Generalisability of the trial results to the predicted population for whom it might be used in the NHS	
1. Is it reasonable to assume that alloHSCT practice is broadly similar in England and Wales to that in other European countries?	We believe it is reasonable to assume that alloHSCT practice is broadly similar in England and Wales as with other European countries, because 52 UK transplant centres are members of the EBMT and work according to EBMT Guidelines (see also response to question 2 below). Please note that the International Coordinating Investigator of the medac-sponsored trial MC-FludT.17/M in children with malignant diseases was Prof. Ajay Vora, Great Ormond Street Hospital for Children, London. Ten patients (14.3%) included in this trial came from UK.
2. Is it plausible that as UK transplant sites are members of the European Society for Blood and	Yes, we believe that UK transplant sites are treating patients according to the EBMT Guidelines. UK transplant physicians are actively involved in the preparation of EBMT guidelines. The

Marrow Transplantation (EBMT), they also treat their patients according to the EBMT Guidelines?	following UK experts were involved in the writing of the 2019 EBMT Handbook for Hematopoietic Stem Cell Transplantation and Cellular Therapies: Persis Amrolia, Jane Apperley, Charles Craddock, Francesco Dazzi, Khaled El-Ghariani, Irina Evseeva, Andrew Gennery, Brenda E. S. Gibson, Diana M. Greenfield, Shelley Hewerdine, Alejandro Madrigal, David I. Marks, Silvia Montoto, John Murray, Stephen Robinson, Vanderson Rocha, Carmen Ruiz de Elvira, Nina Salooja, Basil Sharrack, John A. Snowden, Paul Veys, and Robert Wynn.[25] John Murray is one of 4 Board members of the EBMT, and John Snowdon and Silvia Montoto currently lead 2 of 10 Scientific Council - Working Parties of the EBMT. Furthermore, the compliance of EBMT members with EBMT guidelines is supervised by the Joint Accreditation Committee, International Society for Cell and Gene Therapy (ISCT) and EBMT (JACIE), which offers an inspection-based accreditation process in HSCT against established international standards.
3. Is the population of MC-FludT.14/L Trial II broadly representative of UK clinical practice in terms of age and weight?	Yes, the MC-FludT.14/L study closely reflects routine clinical practice in the UK which is illustrated in Killick et al,[26] the National Cancer Research Institute's protocols for AML (AML15,[27] AML16,[28] AML18[29]) and the 2016 Yorkshire and Humberside Clinical Networks guidelines for AML.[30]
4. In UK clinical practice, how are people at increased risk (not eligible to standard MAC) defined? Is this similar to how they were defined in the clinical trial?	
5. Is it plausible to assume that the time to relapse and time to death observed in the trial for those who received low dose busulfan with fludarabine is generalisable to patients in the UK who receive this regimen (event-free survival at 24 months was 51.2% in the busulfan arm, EFS curves are reported in Appendix 3)?	Yes, it is plausible, although there are only limited published EFS and OS data for BU/FLU RIC from UK. A prospective study in 75 patients (median age 52 years, range 19-68) with MDS receiving alloHSCT with BU/FLU RIC showed 3-year OS of 43% and disease-free survival (DFS) of 41%.[31] This is about 10% lower than respective data from the BU-arm trial of study MC-FludT.14/L Trial II (56.3% / 49.7%).[1] In a subsequent study of the UK group, the BU/FLU regimen was used to treat 192 patients with MDS-AML/MDS. Median age of patients was 57 years (range 21-72). 3-year OS/EFS estimates were 50%/38%.[32]
6. Which patient's characteristics would be	The most important treatment-effect modifiers are age, disease status at transplant, risk group,

expected to be treatment-effect modifiers?	donor type (MRD or MUD), pre-treatment, and comorbidity index. Please note that subgroup analysis of patients in the pivotal MC-FludT.14/L Trial II analysed treatment outcome in such subgroups and revealed better EFS and OS in nearly all subgroups (see respective forest plots for EFS and OS in the clinical study report[1]).
Issue 5: Limited evidence of comparative effectiveness of treosulfan versus other conditioning regimens particularly myeloablative conditioning (MAC) regimens	
1. Would clinicians consider that the evidence from MC-FludT.14/L Trial II is good enough for treosulfan and fludarabine to be used in patients who can tolerate MAC? Or would patients who can tolerate MAC receive MAC regimen anyway?	Yes, we believe that clinicians would consider the evidence sufficient to utilise TREO/FLU in MAC eligible patients. We want to underline that the TREO/FLU regimen is a MAC regimen albeit with reduced toxicity compared to other MAC regimens. TREO/FLU is therefore also suited for younger patients without comorbidities. It is important that clinicians should have the choice of using TREO/FLU in patients who could tolerate MAC according to their assessment of individual patient characteristics. In patients with AML and MDS, the efficacy and safety of TREO/FLU has been confirmed in the MC-FludT.7/AML[3] and MC-FludT.8/MDS[4] trials, many published investigator-initiated trials),[2,6,8–11,33] and data from about 1,000 patients documented in the EBMT registry.[6,34] Currently, the TREO/FLU regimen is already used by many transplant centres in Europe, USA, and other countries.
2. Should treosulfan and fludarabine be compared with MAC regimens?	There are published retrospective studies from the EBMT registry that compared TREO/FLU with other MAC regimens which demonstrated that TREO-based conditioning is at least as good as other MAC regimens. For example, Shimoni et al. performed a retrospective analysis of all alloHSCTs for MDS performed between 2000 and 2011 and reported to the EBMT (n = 2,516).[34] They identified 480 patients given TREO/FLU and compared their outcomes to patients given various MAC (n = 1,090) and RIC (n = 946) regimens. The authors concluded that TREO/FLU is associated with similar low relapse rates as other MAC regimens and similar low NRM as RIC regimens, resulting in improved outcome over both RIC and MAC. Therefore, TREO/FLU might be the preferred regimen for alloHSCT in MDS.
3. Could the EBMT and Centre for International Blood and Marrow Transplant Research (CIBMTR) analyses be improved (for example by	The EBMT analysis already included a matched pair analysis (using propensity score matching methods) in which patients from the registry receiving FLU/MEL or BU/CY conditioning were matched 1:1 with TREO patients from MC-FludT.14/L Trial II for the most important confounding

adjusting for age, matching and incorporating clinical opinion to identify the clinically significant variables to adjust for) while noting the limitations stated by the ERG, be incorporated into the CE model as a scenario to provide more robust/improved evidence of comparative effectiveness of treosulfan versus other conditioning regimens?	factors. The CIBMTR analysis is based on patients who were selected according to the inclusion/exclusion criteria of medac's pivotal MC-FludT.14/L Trial II. In the CIBMTR analysis, EFS/OS results with BU/FLU RIC were comparable to the BU/FLU arm in medac's MC-FludT.14/L Trial II, which shows that the treatment results with BU/FLU in the medac trial are representative and were not underestimated. Since for children the same BU/FLU regimen as in adults is frequently used, and the available evidence is less robust than that for adults, we decided not to perform an additional cost-effectiveness analysis.
4. Do the differences between the trials included in the indirect treatment comparison (ITC) justify the company's decision not to conduct an ITC?	Studies published with other conditioning regimens are difficult to compare to the results of the medac trial with TREO/FLU conditioning because the patient characteristics are largely different. Please note that there is no published prospective alloHSCT study in a representative number of patients available which showed better EFS, OS, and NRM in adult AML/MDS than that seen with TREO/FLU conditioning in study MC-FludT.14/L Trial II.
5. Are these patients' characteristics of the trial expected to be treatment-effect modifiers for example, number of patients receiving matched unrelated alloHSCT and age?	Yes, the patient characteristics of the trial are expected to be treatment-effect modifiers. Please note that the advantage of TREO-based conditioning has been shown in matched unrelated donors (MUD) as well as matched related donors (MRD) and in patients of different age groups (< or > 50 years of age).
Issue 6: Mortality modelling	
1. If patients did not relapse after transplantation, would they be considered cured at 1 year, 2 years, 5 years?	There is general consensus among transplant experts that patients surviving alloHSCT disease-free for more than 5 years can be expected to be cured . Relapse rates and death after 5 years are rare. Most survival curves in such patients show a plateau at 5 years. In the medac MC-FludT.14/L Trial II, the plateau was reached after 40 months. A 5-year functional cure point was also suggested by the two clinical experts interviewed for validation of the economic analysis, both of whom suggested that this option would be more appropriate than the alternative fixed cure points at 1-4 years or allowing the cure point to be defined when the adjusted life table mortality exceed mortality from the OS curve.

Appendix 1

Efficacy results of MC-FludT.14/L Trial II (from ERG report, Table 4.7)

	Treosulfan (10 g/m²/day) + Fludarabine (30 g/m²/day)	Busulfan (3.2 mg/kg/day) + Fludarabine (30 g/m²/day)
Number randomised	280	290
Number analysed (FAS: patients who received conditioning treatment and HSCT)	268	283
Median follow-up[a], months (range of those surviving)	29.7 (3.0 to 52.1)	29.4 (3.0 to 54.3)
Primary outcome – Event-free survival (EFS) within 24 months after alloHSCT		
Patients with event	97 (36.2%)	137 (48.4%)
Death[b]	35 (13.1%)	56 (19.8%)
Relapse/progression[b]	61 (22.8%)	72 (25.4%)
Primary graft failure[b]	1 (0.4%)	1 (0.4%)
Secondary graft failure[b]	0 (0.0%)	8 (2.8%)
Event-free survival at 12 months[c] [%] (95% CI)	70.0 (64.1 to 75.1)	60.8 (54.9 to 66.3)
Event-free survival at 24 months[c] [%] (95% CI)	65.7 (59.5 to 71.2)	51.2 (45.0 to 57.0)
Event-free survival at 36 months[c] [%] (95% CI)	[redacted]	[redacted]
Hazard ratio [HR][d] (95% CI)	0.64 (0.49 to 0.84)	
Secondary outcome – Overall survival (OS) within 24 months after alloHSCT		
Patients with event	81 (30.2%)	112 (39.6%)
Overall survival at 12 months[c] [%] (95% CI)	77.8 (72.3 to 82.3)	71.8 (66.1 to 76.7)
Overall survival at 24 months[c] [%] (95% CI)	72.7 (66.8 to 77.8)	60.2 (54.0 to 65.8)
Overall survival at 36 months[c] [%] (95% CI)	[redacted]	[redacted]
HR[d] (95% CI)	0.64 (0.48 to 0.87)	

	Treosulfan (10 g/m²/day) + Fludarabine (30 g/m²/day)	Busulfan (3.2 mg/kg/day) + Fludarabine (30 g/m²/day)
Secondary outcome – Cumulative incidence of relapse/progression 24 months after alloHSCT		
Patients with event	61 (22.8%)	72 (25.4%)
Patients without event (censored) or with competing event	207 (77.2%)	211 (74.6%)
Censored	171 (63.8%)	146 (51.6%)
Death[b]	35 (13.1%)	56 (19.8%)
Primary graft failure[b]	1 (0.4%)	1 (0.4%)
Secondary graft failure[b]	0 (0.0%)	8 (2.8%)
Cumulative incidence at 12 months [%] (95% CI)	19.1 (14.4 to 23.8)	21.7 (16.9 to 26.5)
Cumulative incidence at 24 months [%] (95% CI)	22.0 (16.9 to 27.1)	25.2 (20.0 to 30.3)
Cumulative incidence at 36 months [%] (95% CI)	[redacted]	[redacted]
HR[e] (95% CI)	0.82 (0.59 to 1.16)	
Secondary outcome - engraftment		
Primary graft failure[f]	1/268 (0.4%)	1/283 (0.4%)
Secondary graft failure[f]	0/263 (0.0%)	8/279 (2.9%)
Secondary outcome (not specified in scope) – Cumulative incidence of non-relapse mortality (NRM) 24 months after alloHSCT		
Patients with event	35 (13.1%)	56 (19.8%)
Patients without event (censored) or with competing event	233 (86.9%)	227 (80.2%)
Censored	171 (63.8%)	146 (51.6%)
Relapse/Progression[b]	61 (22.8%)	72 (25.4%)
Primary Graft Failure[b]	1 (0.4%)	1 (0.4%)
Secondary Graft Failure[b]	0 (0.0%)	8 (2.8%)
Cumulative incidence at 12 months [%] (95% CI)	10.5 (6.8 to 14.2)	14.3 (10.2 to 18.4)

	Treosulfan (10 g/m²/day) + Fludarabine (30 g/m²/day)	Busulfan (3.2 mg/kg/day) + Fludarabine (30 g/m²/day)
Cumulative incidence at 24 months [%] (95% CI)	12.0 (8.0 to 15.9)	20.4 (15.5 to 25.2)
Cumulative incidence at 36 months [%] (95% CI)	[redacted]	[redacted]
HR[g] (95% CI)	0.63 (0.41 to 0.97)	
Secondary outcome (not specified in scope) – Transplantation-related mortality (TRM)		
Patients with event	33 (12.3%)	58 (20.5%)
Patients without event	235 (87.7%)	225 (79.5%)
Transplantation-related mortality at 12 months[c] [%] (95% CI)	11.7 (8.3 to 16.3)	16.2 (12.2 to 21.3)
Transplantation-related mortality at 24 months[c] [%] (95% CI)	12.8 (9.2 to 17.7)	24.1 (19.1 to 30.2)
HR[d] (95% CI)	0.52 (0.34 to 0.82)	

Based on Tables 18, 19, 20, 21 and 23 of the CS[1]

[a] Based on reverse Kaplan-Meier estimates for overall survival; [b] Only if this event occurred first; [c] Based on Kaplan-Meier estimates; [d] Adjusted for donor type as factor, and risk group and centre as strata using Cox regression model; [e] Adjusted for donor type as factor and risk group as stratum using Fine and Gray model; [f] Rate of primary/secondary graft failure calculated as number of patients with graft failure by the number of patients at risk; [g] Adjusted for donor type as factor

alloHSCT = allogeneic haematopoietic stem cell transplantation; CI = confidence interval; CS = company submission; EFS = event-free survival; FAS = full analysis set; HR = hazard ratio; HSCT = Haematopoietic stem cell transplantation; NRM = non-relapse mortality; OS = Overall survival; TRM = transplantation-related mortality

Appendix 2

Patients characteristics from MC-FludT.14/L Trial II – pooled AML and MDS patients (from company's submission, Table 36)

Variable	Busulfan	Treosulfan	Total
N	283	268	551
N (male)	173	162	335
N (female)	110	106	216
Sex (male)	61.13%	60.45%	60.80%
Sex (female)	38.87%	39.55%	39.20%
Age (mean)	59.9	59.3	59.6
Age (SD)	6.0	6.5	6.3
Weight (mean), kg	79.4	80.9	80.2
Weight (SD), kg	17.7	16.7	17.3
n (MRD)	68	62	130
n (MUD)	215	206	421
% (MRD)	24.03%	23.13%	23.59%
% (MUD)	75.97%	76.87%	76.41%
BSA (mean), m^2	1.921	1.942	1.931
BSA (SD), m^2	0.241	0.227	0.235
RBC transfusion dependency, n (No)	219	216	435
RBC transfusion dependency, n (Yes)	64	52	116
RBC transfusion dependency, % (No)	77.39%	80.60%	78.95%
RBC transfusion dependency, % (Yes)	22.61%	19.40%	21.05%

Appendix 3

Table 1 Kaplan Meier estimates of EFS - MC-FludT.14/L Trial II

Source: Company's submission Figure 4

APPENDIX 4

Update of results of trial MC-FludT.17/M (Executed: 09-Jan-2019)

Summary results of transplantation-related mortality (Full Analysis Set)	Treosulfan (N = 70)
Subjects with event	[redacted]
Subjects without event	[redacted]
Transplantation-related mortality at 100 days[a] [%] 90% CI	[redacted]
Transplantation-related mortality at 12 months[a] [%] 90% CI	[redacted]
Transplantation-related mortality at 24 months[a] [%] 90% CI	[redacted]
Transplantation-related mortality at 36 months[a] [%] 90% CI	[redacted]
[a] Based on Kaplan-Meier estimates [Table 14.2.2A: Program: EBMT 2019 /SurvivalTRM /t_trm_fas]	

Summary results of overall survival (Full Analysis Set)	Treosulfan (N = 70)
Median follow-up[a] [months] (range of those surviving)	[redacted]
Subjects with event	[redacted]
Subjects without event	[redacted]
Overall survival at 12 months[b] [%] 90% CI	[redacted]
Overall survival at 24 months[b] [%] 90% CI	[redacted]
Overall survival at 36 months[b] [%] 90% CI	[redacted]
[a] Based on reverse Kaplan-Meier estimates [b] Based on Kaplan-Meier estimates [Table 14.2.3A: Program: EBMT 2019 /SurvivalOS_RFPFS_EFS /t_os_fas]	

Summary results of relapse/progression (Full Analysis Set)	Treosulfan (N = 70)
Subjects with event	[redacted]
Subjects without event Censored Death[a] Primary graft failure[a] Secondary graft failure[a]	[redacted]
Cumulative incidence of relapse/progression at 12 months [%] 90% CI	[redacted]
Cumulative incidence of relapse/progression at 24 months [%] 90% CI	[redacted]

Summary results of relapse/progression (Full Analysis Set)	Treosulfan (N = 70)
Cumulative incidence of relapse/progression at 36 months [%] 90% CI	[redacted]

[a] Only if this event occurred first

[Table 14.2.4A: Program: EBMT 2019 /Survival RPS_NRM /t_cuminc_rps_summary_fas]

References

1. Medac. *MC-FludT.14/L Trial II - Clinical Study Report*. (2018).
2. Gran, C. *et al.* Treosulfan Conditioning for Allogeneic Transplantation in Multiple Myeloma Improved Overall Survival in Upfront Hematopoietic Stem Cell Transplantation — a Large Retrospective Study By the Chronic Malignancies Working Party of the EBMT. *Blood* **132**, 3464 (2018).
3. Medac. *MC-FludT.7/AML - Clinical Study Report*. (2011).
4. Medac. *MC-FludT.8/MDS - Clinical Study Report*. (2010).
5. Deeg, H. J. *et al.* Transplant Conditioning with Treosulfan/Fludarabine with or without Total Body Irradiation: A Randomized Phase II Trial in Patients with Myelodysplastic Syndrome and Acute Myeloid Leukemia. *Biol. Blood Marrow Transplant.* (2018). doi:10.1016/j.bbmt.2017.12.785
6. Nagler, A. *et al.* Long-term outcome after a treosulfan-based conditioning regimen for patients with acute myeloid leukemia: A report from the Acute Leukemia Working Party of the European Society for Blood and Marrow Transplantation. *Cancer* **123**, 2671–2679 (2017).
7. Yerushalmi, R. *et al.* Fludarabine and treosulfan compared with other reduced-intensity conditioning regimens for allogeneic stem cell transplantation in patients with lymphoid malignancies. *Bone Marrow Transplant.* **50**, 1526–1535 (2015).
8. Gyurkocza, B. *et al.* Treosulfan, fludarabine, and 2-Gy total body irradiation followed by allogeneic hematopoietic cell transplantation in patients with myelodysplastic syndrome and acute myeloid leukemia. *Biol. Blood Marrow Transplant.* **20**, 549–555 (2014).
9. Schmitt, M. *et al.* Conditioning with treosulfan and fludarabine for patients with refractory or relapsed non-Hodgkin lymphoma. *Mol. Clin. Oncol.* **2**, 773–782 (2014).
10. Michallet, M. *et al.* Phase II prospective study of treosulfan-based reduced-intensity conditioning in allogeneic HSCT for hematological malignancies from 10/10 HLA-identical unrelated donor. *Ann. Hematol.* **91**, 1289–1297 (2012).
11. Casper, J. *et al.* Allogeneic hematopoietic SCT in patients with AML following treosulfan/fludarabine conditioning. *Bone Marrow Transplant.* **47**, 1171–1177 (2012).
12. Boztug, H. *et al.* European Society for Blood and Marrow Transplantation Analysis of Treosulfan Conditioning Before Hematopoietic Stem Cell Transplantation in Children and Adolescents With Hematological Malignancies. *Pediatr. Blood Cancer* **63**, 139–48 (2016).
13. Boztug, H. *et al.* Treosulfan-based conditioning regimens for allogeneic HSCT in children with acute lymphoblastic leukaemia. *Ann. Hematol.* **94**, 297–306 (2014).
14. Nemecek, E. R. *et al.* Treosulfan, Fludarabine, and Low-Dose Total Body Irradiation for Children and Young Adults with Acute Myeloid Leukemia or Myelodysplastic Syndrome Undergoing Allogeneic Hematopoietic Cell Transplantation: Prospective Phase II Trial of the Pediatric Blood . *Biol. Blood Marrow Transplant.* **24**, 1651–1656 (2018).
15. Maschan, M. *et al.* TCR-alpha/beta and CD19 depletion and treosulfan-based conditioning regimen in unrelated and haploidentical transplantation in children with acute myeloid leukemia. *Bone Marrow Transplant.* **51**, 668–674 (2016).
16. Wachowiak, J. *et al.* Treosulfan-based preparative regimens for allo-HSCT in

childhood hematological malignancies: A retrospective study on behalf of the EBMT pediatric diseases working party. *Bone Marrow Transplant.* **46**, 1510–1518 (2011).
17. Peters, C., Dallisier, A. & Beohou, E. Retrospective EBMT-PD / IE-WP analysis: Treosulfan or Busulfan based conditioning before allogeneic haematopoietic stem cell transplantation in paediatric patients with non-malignant diseases. (2017).
18. Peters, C. *Meta-analysis on Treosulfan for conditioning in children and adolescents before haematopoietic stem cell transplantation: Stastical results [AIC]*. (2018).
19. EWOG-MDS. Guidelines for Hematopoietic Stem Cell Transplantation (HSCT) in Childhood MDS and JMML for Patients enrolled in EWOG-MDS Studies. Version 1.3. in *EWOG-MDS Consensus Conference Freiburg* 1–19 (2016).
20. European Society for Blood and Marrow Transplantation & European Society for Immunodeficiencies. *EBMT / ESID Guidelines for Haematopoietic Stem Cell Transplantation for Primary Immunodeficiencies*. *ESID EBMT HSCT Guidelines 2017* (2017).
21. Scott, B. L. *et al.* Myeloablative versus reduced-intensity hematopoietic cell transplantation for acute myeloid leukemia and myelodysplastic syndromes. *J. Clin. Oncol.* **35**, 1154–1161 (2017).
22. Baron, F. *et al.* Reduced-intensity conditioning with fludarabine and busulfan versus fludarabine and melphalan for patients with acute myeloid leukemia: A report from the Acute Leukemia Working Party of the European Group for Blood and Marrow Transplantation. *Cancer* **121**, 1048–1055 (2015).
23. Center for International Blood and Marrow Transplant Research (CIBMTR). *Report of the comparison of conditioning regimens against medac data*. (2019).
24. European Society for Blood and Marrow Transplantation (EBMT), Iacobelli, S., Koster, L. & Biezen, A. Van. *Re-analysis of EBMT-registry data on Fludarabine/Melphalan and Busulfan/Cyclophosphamide based conditioning treatment compared to Fludarabine/Treosulfan based conditioning of adult AML and MDS patients treated in MC-FludT.14/L phase III by matched pairs.* . (2019).
25. European Society for Blood and Marrow Transplantation (EBMT). *The EBMT Handbook Hematopoietic Stem Cell Transplantation and Cellular Therapies*. (2019).
26. Killick, S. B. *et al.* Guidelines for the diagnosis and management of adult myelodysplastic syndromes. *Br. J. Haematol.* **164**, 503–525 (2014).
27. National Institute Health Research (NIHR). AML15: Protcol for patients aged under 60. (2004).
28. National Institute Health Research (NIHR). AML16: A trial for older patients with acute AML and high risk myelodysplastic syndrome. 1–46 (2006).
29. National Institute for Health Research (NIHR). AML18: Risk Myelodysplastic Syndrome - Protocol under development. (2019). Available at: https://ukctg.nihr.ac.uk/trials/trial-details/trial-details?trialNumber=ISRCTN78449203.
30. Yorkshire and Humberside Clinical Networks. *Clinical Guidelines for Leukaemia and other Myeloid Disorders – Myeloproliferative Neoplasms*. (2016).

31. Lim, Z. Y. *et al.* Outcomes of alemtuzumab-based reduced intensity conditioning stem cell transplantation using unrelated donors for myelodysplastic syndromes. *Br. J. Haematol.* **135**, 201–9 (2006).
32. Potter, V. T. *et al.* Long-Term Outcomes of Alemtuzumab-Based Reduced-Intensity Conditioned Hematopoietic Stem Cell Transplantation for Myelodysplastic Syndrome and Acute Myelogenous Leukemia Secondary to Myelodysplastic Syndrome. *Biol. Blood Marrow Transplant.* **20**, 111–117 (2014).
33. Eapen, M. *et al.* Hematopoietic cell transplant for acute myeloid leukemia and myelodysplastic syndrome: conditioning regimen intensity. *Blood Adv.* **2**, 2095–2103 (2018).
34. Shimoni, A. *et al.* Fludarabine and treosulfan conditioning is associated with a more favorable outcome after allogeneic stem cell transplantation in myelodysplastic syndrome. A survey on behalf of the Chronic Malignancies Working Party of EBMT.No Title. in *41st Annual Meeting of the EBMT* (2015).

NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE

Technical report

Treosulfan with fludarabine for malignant disease before allogeneic stem cell transplant

This document is the post-engagement technical report for this appraisal. It has been prepared by the technical team with input from the lead team and chair of the appraisal committee.

The technical report and stakeholder's responses to it are used by the appraisal committee to help it make decisions at the appraisal committee meeting. Usually, only unresolved or uncertain key issues will be discussed at the appraisal committee meeting. A draft version of this technical report was sent out for consultation between September 12th and October 10th. The draft report included a list of issue that have an impact on the uncertainty of the company's estimates of clinical or cost-effectiveness. The aim of the consultation was to seek feedback from consultees and commentators on these issues to help inform the technical team's favoured modelling assumptions.

The aim of the post-engagement version of the technical report is to:

- Summarise the feedback that was received on the issues that were identified originally
- Explain how the feedback has or has not been helpful in resolving areas of uncertainty

Usually, only unresolved or uncertain key issues will be discussed at the appraisal committee meeting.

Issue date: November 2019

The technical report includes:

- topic background based on the company's submission
- a commentary on the evidence received and written statements
- technical judgements on the evidence by the technical team
- reflections on NICE's structured decision-making framework.

This report is based on:

- the evidence and views submitted by the company, consultees and their nominated clinical experts and patient experts and
- the evidence review group (ERG) report.

The technical report should be read with the full supporting documents for this appraisal.

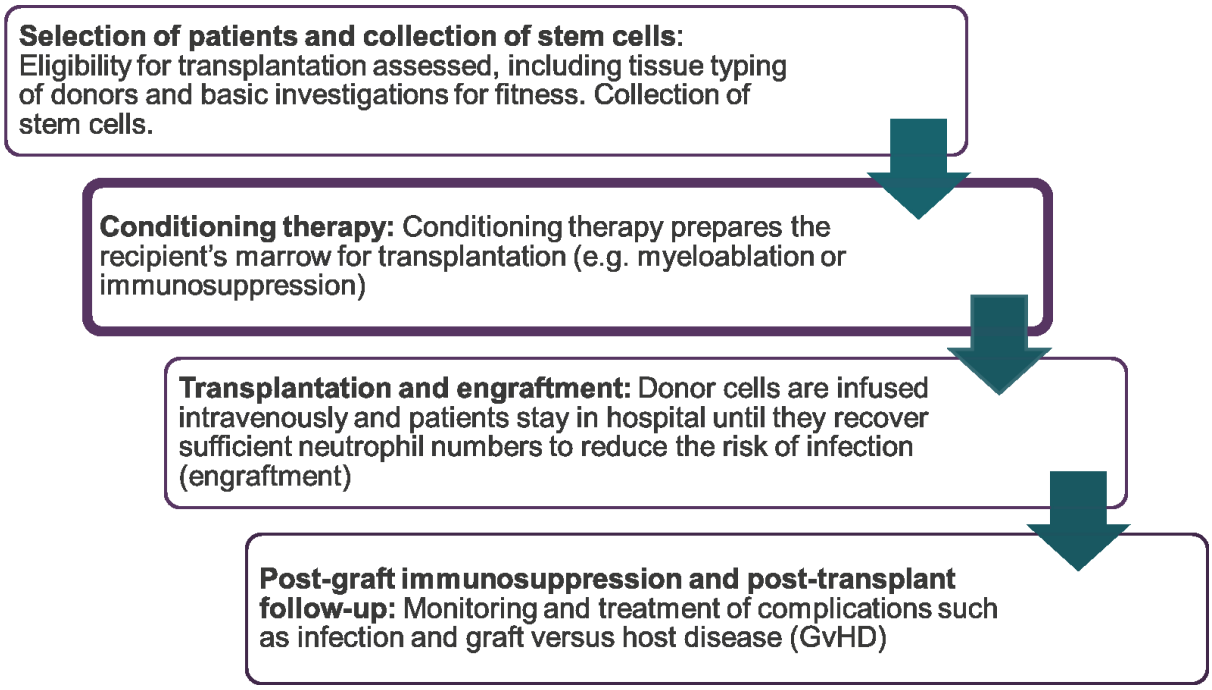
1. Topic background

1.1 Disease background

- Haematopoietic stem cell transplant (HSCT) is a potentially curative therapy for more than 70 malignant diseases.
- In the UK, the most common malignant indications for allogeneic HSCT (alloHSCT) are acute myeloid leukaemia (AML; 36%), acute lymphoblastic leukaemia (ALL; 16%) and myelodysplastic syndrome (MDS) or myelodysplastic/myeloproliferative neoplasms (MDS/MPN: 13%).
- Generally, alloHSCT is recommended to be performed at early stages of the disease, as soon as complete remission is achieved in patients with high-risk diseases.
- Remission status at the time of HSCT is an important prognostic factor in the risk of relapse.
- Before undergoing HSCT, patients receive a **conditioning regimen** in order to prepare the patient's marrow for transplantation. Conditioning treatments are usually chemotherapy alone or chemotherapy associated with radiotherapy.

1.2 Treatment pathway

1.2.1 HSCT

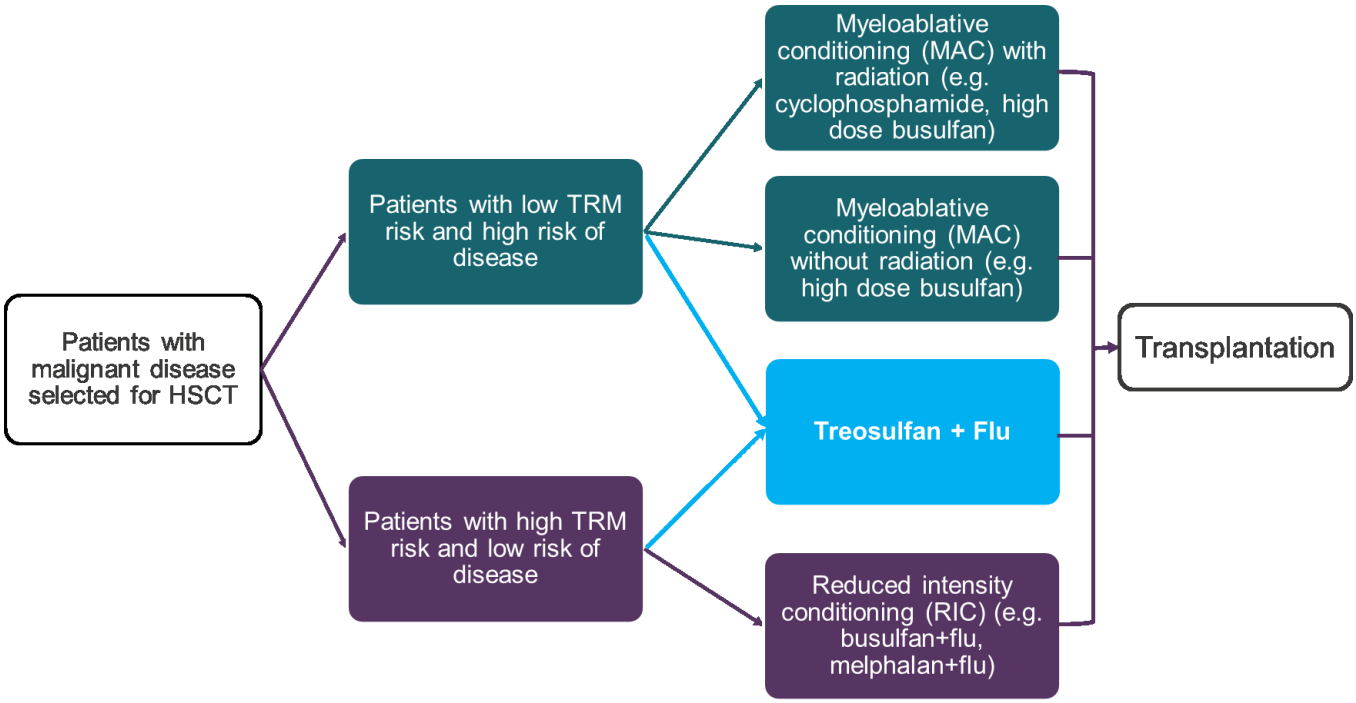


1.2.2 Conditioning treatments

- High intensity myeloablative conditioning (MAC) and reduced intensity conditioning (RIC) are 2 types of conditioning regimens.[1]
  - MAC regimens use total body irradiation and/or high-dose alkylating agents to cause irreversible pancytopenia, thereby minimising the risk of disease recurrence.
  - RIC use lower doses of total body irradiation or alkylating agents than MAC regimens, resulting in cytopenia which may not be irreversible.
- The company proposes that the licensed treosulfan regimen is a reduced-toxicity regimen, which is myeloablative, but has lower toxicity than standard MAC regimens.

[1] Bacigalupo A, et al, Defining the intensity of conditioning regimens: working definitions. Biol Blood Marrow Transplant. 2009 Dec;15(12):1628-33.

- The 3 aims of a conditioning treatment are:
  - To reduce the tumour burden when the disease is neoplastic
  - To eliminate the self-renewing capacity of the patient's own haematopoiesis
  - To suppress the recipient's immune system in order to allow engraftment of stem cells
- The European Society for Blood and Marrow Transplantation (EBMT) recommends that patients with high risk of transplantation related mortality (TRM) and a low disease risk should receive a different conditioning regimen from patients with low risk of TRM and high-risk disease



Ref: Company's submission, document B figure 2

1.3 Decision problem

	Final scope issued by NICE	Evidence used in the model
Population	Adults, children and young people with malignant disease that is in remission before alloHSCT	Adults with AML and MDS that is in remission before alloHSCT
Intervention	Treosulfan with fludarabine	As final scope
Comparators	Standard myeloablative regimens - Cyclophosphamide and total body irradiation - Cyclophosphamide and busulfan - Cyclophosphamide and thiotepa - High dose busulfan with fludarabine with or without thiotepa Reduced intensity regimens - Low dose busulfan with fludarabine - Melphalan plus fludarabine	Low dose busulfan with fludarabine
Outcomes	The outcomes measures include: - Overall survival - Event-free survival - Rates of relapse - Success of transplantation (engraftment) - Adverse effects of treatments - Health-related quality of life	Additional including: - non-relapse mortality (NRM)

1.4 Clinical evidence

	MC-FludT.14/L Trial II	
Population	Adult patients with AML or MDS who are at increased risk for standard conditioning therapies (not eligible for standard myeloablative conditioning busulfan- or TBI-based regimens). Patients at increased risk for standard conditioning therapies were defined as ≥50 years and/or HCT-CI score≥2	
	Treosulfan + flud. (N=268)	Busulfan + flud. (N=283)
Primary endpoint: event-free survival (EFS)		
Median follow-up (months)	29.7	29.4
Death	35 (13.1%)	56 (19.1%)
Relapse/Progression	61 (22.8%)	72 (25.4%)
Primary graft failure	1 (0.4%)	1 (0.4%)
Secondary graft failure	0 (0.0%)	8 (2.8%)
Event-free survival at 24 months (95%CI)	65.7% (59.5, 71.2)	51.2% (45.0, 57.0)
Hazard ratio (95% CI)	0.64 (0.49, 0.84)	
Secondary endpoint		
Overall survival at 24 months (%) (95% CI)	72.7 (66.8, 77,8)	60.2 (54.0, 65.8)
Hazard ratio (95% CI)	0.64 (0.48, 0.87)	
Transplantation related mortality at 24 months (%) (95% CI)	12.8 (9.2 to 17.7)	24.1 (19.1 to 30.2)
Hazard ratio (95% CI)	0.52 (0.34 to 0.82)	
Cumulative incidence of non relapse mortality at 24 months (%) (95% CI)	12.0 (8.0 to 15.9)	20.4 (15.5 to 25.2)
Hazard ratio (95% CI)	0.63 (0.41, 0.97)	

- Protocol amendment 03: the treosulfan dose was reduced from 3x14mg/m^2 to 3x10mg/m^2, as partly unfavourable findings regarding increased infectious complications after Treosulfan treatment were associated with an imbalanced dosing.

1.5 Model structure

- Partitioned survival model
- Patients enter the model having received HSCT and start in the induction/HSCT health state
- Subsequent health state occupancy is determined by EFS and OS curves

1.6 Key model assumptions

Population	• People with malignant disease that is in remission before alloHSCT • Evidence is based on patients from the MC-FludT.14/L Trial II only– adults with AML or MDS that is in remission before alloHSCT, at increased risk for standard conditioning therapy (that is, not eligible for standard high-intensity MAC)
Intervention	Treosulfan (10 mg/m^2) x3 doses and fludarabine (30 mg/m^2) x 5 doses
Comparators	Busulfan (3.2 mg/kg) x3 doses and fludarabine (30 mg/m^2) x 5 doses
Concomitant treatments	The same therapies were used concomitantly with treosulfan and busulfan except for phenytoin
Mortality modelling and cure point	• If patients did not relapse 5 years after transplantation, they are considered cured • EFS and OS curves are used until the cure point, then HSCT-specific life tables are used to model mortality.
Health-related quality-of-life	• Utility values based on literature – Clinical trial did not collect HRQoL data
Costs and resource use	• Administration costs are excluded from calculation • Wastage costs are applied, with 100% wastage vial wastage assumed in the base case • All costs associated with adverse events management were incurred in an inpatient setting

2. Summary of the technical report

After technical engagement, the technical team has collated the comments received and, if relevant, updated the judgement made by the technical ream and rationale. The issues that were considered at technical engagement are described in detail in section 3 below, along with the feedback that was received. The following table summarises the current status of each issue in terms of the technical team's view on the level of outstanding uncertainty.

Issue identified pre-engagement and issue status following engagement	**Response to consultation**	**Technical team judgement after engagement**
Issue 1 – How does treosulfan differ from other alkylating agents, such as busulfan, in terms of its mechanism of action and pharmacokinetics? How does this translate into clinical benefit? The company states that non relapse mortality is the main contributor to the overall survival benefit, but this is included in the composite outcome of event-free survival (EFS) as event free death and differences in relapse rates appear modest. Clarification is requested on to what extent this treatment prevents death related to the transplant and post-transplant complications, and to what extent it reduces the risk of relapse and subsequent death from malignancy **For discussion**	The main advantage of treosulfan is the lower toxicity compared with other regimens, especially the busulfan and fludarabine regimen. This leads to lower non-relapse mortality.	The technical team understand that the main benefit of treosulfan is reduced toxicity leading to lower non-relapse mortality versus a busulfan and fludarabine regimen. Patients are less likely to die from the transplant or associated infections, graft vs host disease. It is in unclear whether treosulfan is a myeloablative regimen at the dose administered in the trial and specified in the marketing authorisation.
Issue 2 – To what extent can the results of the trial be extrapolated to a broader population? The population included in the submission is narrower than the population of the scope and the marketing authorisation. Treosulfan in combination with fludarabine is	According to the company, the trial results are generalisable to a broader population, because the main benefit of treosulfan is the reduced toxicity which is not impacted by the underlying disease.	The technical team considers that if the main benefit of treosulfan is on the toxicity and not on disease relapse, then it is plausible to assume that the benefits of treosulfan-versus busulfan-based

Issue identified pre-engagement and issue status following engagement	Response to consultation	Technical team judgement after engagement
indicated as part of conditioning treatment prior to alloHSCT in adult patients with malignant and non-malignant diseases, and in paediatric patients older than one month with malignant diseases. This appraisal focuses on malignant diseases. The MC-FludT.14/L Trial II included adult patients with acute myeloid leukaemia (AML) or myelodysplastic syndrome (MDS) who are at increased risk for standard conditioning therapies (that is, not eligible for standard myeloablative conditioning [MAC] regimens based on busulfan- or total body irradiation [TBI]). The company believes that the evidence from MC-FludT.14/L Trial II is applicable to a broader population of adults, children and young people with malignant disease that is in remission before allogeneic hematopoietic stem cell transplant (alloHSCT), in line with the final scope. However, there is limited evidence for people with malignancies other than AML and MDS and adults who would be eligible for standard myeloablative regimens. Clinical opinion is sought on whether the efficacy of treosulfan can be broadened to a wider population of people with malignant disease requiring myeloablative conditioning treatment, including children and patients who can tolerate standard myeloablative regimens. **For discussion**	Evidence has been provided within the large studies in patients with other malignancies.	regimens are generalisable to a broader population of adults with other malignant diseases. Regarding children, the company refers to the MC-FludT.17/M trial, the EBMT registry data and a meta-analysis to support the efficacy of treosulfan in children. However, no cost-effectiveness analysis is available.

Issue identified pre-engagement and issue status following engagement	Response to consultation	Technical team judgement after engagement
Issue 3 – The submission focuses on a low-intensity regimen In the cost-effectiveness analysis, the only comparator included was low-dose busulfan with fludarabine, which is the comparator in MC-FludT.14/L Trial II and does not include other comparators from the scope. The ERG considers that the evidence submitted is sufficient to support an assessment of the cost effectiveness of the licensed dose of treosulfan (3 doses of 10mg/m^2) in combination with fludarabine versus reduced-intensity conditioning (RIC) busulfan and fludarabine in adults with AML or MDS, who are at increased risk for standard conditioning therapies (as explained in Issue 2). But is not sufficient to support an assessment of treosulfan versus any of the other comparators defined in the scope, including standard myeloablative regimens such as high dose busulfan and fludarabine, cyclophosphamide and total body irradiation, cyclophosphamide and busulfan. **For discussion**	Busulfan is the commonest regimen used in the UK, especially low dose busulfan because most patients are older and cannot tolerate high doses. It is a good proxy for conditioning regimens in the UK. Patients who can tolerate a standard myeloablative regimen are unlikely to have a RIC regimen which would increase the risk of relapse.	Busulfan is the most common alkylating agent used in reduced intensity conditioning regimens the UK and the regimen used in FludT.14/L Trial II and is a good proxy for reduced intensity conditioning regimens used in the NHS. There is no evidence on which to assess comparative efficacy of the licensed dose of treosulfan in adults who would otherwise receive standard myeloablative regimens.

Issue identified pre-engagement and issue status following engagement	Response to consultation	Technical team judgement after engagement
Issue 4 – Generalisability of the trial results to the predicted population for whom it might be used in the NHS (that is, people who would otherwise get reduced intensity busulfan) Patients included in the MC-FludT.14/L trial were adults with AML or MDS who are at increased risk for standard conditioning therapies (that is, not eligible for standard-dose busulfan- or total body irradiation (TBI)-based regimens). Increased risk for standard conditioning therapies was defined as aged ≥50 years at transplant and/or hematopoietic cell transplantation co-morbidity index (HCT-CI) score≥2. It is not clear how this definition relates to the definition of increased risk generally applied in UK clinical practice and whether clinical practice for transplant in the UK is similar to the trial. Patients included in the trial were mainly from Germany and no UK patients were included. Clinical advice would be valued on whether the population of MC-FludT.14/L Trial II is broadly representative of UK clinical practice. **Agreed**	The company states that the MC-FludT.14/L study reflects the routine clinical practice in the UK, and it is plausible to assume that the time to relapse and death observed in the trial is generalisable to patients in the UK who receive this regimen. However, there are only limited published EFS and OS data for busulfan and fludarabine RIC from the UK.	The technical team agree that it is reasonable to assume that alloHSCT practice is similar in England in Wales as with other European countries. The results of the trial, in terms of the short-term primary and secondary endpoints, are generalisable to the NHS population, but overall survival data are immature.

Issue identified pre-engagement and issue status following engagement	Response to consultation	Technical team judgement after engagement
Issue 5 – Limited evidence of comparative effectiveness treosulfan versus other conditioning regimens particularly standard, high-intensity MAC regimens Although some indirect treatment comparisons were feasible, the company did not perform and did not include any in the cost effectiveness modelling because the company believes that there are differences in the trial and registry population which may affect the validity of the ITC. The technical team's preference is that additional analyses providing more indirect comparison with other conditioning regimens should be performed. The technical team seeks clinical opinion on whether the differences between trials included in the ITC justify the company's approach and whether the registry studies could be improved. The technical team believe that the evidence submitted does not support the use of treosulfan in people who can tolerate higher intensity myeloablative regimens. **For discussion**	The company states that the EBMT analysis already included a matched pair analysis using propensity score matching methods. The CIBMTR analysis is based on patients who were selected according to the inclusion criteria of the MC-FludT.14/L trial. It is difficult to compare studies published with other conditioning regimen to the results of the MC-FludT.14/L trial results because patient's characteristics were largely different. Patient's characteristics in the trial are expected to be treatment-effect modifiers	The technical team believe that the evidence submitted does not support the use of treosulfan in people who can tolerate standard, higher-intensity MAC regimens.

Issue identified pre-engagement and issue status following engagement	Response to consultation	Technical team judgement after engagement
Issue 6 – Mortality modelling Mortality is modelled using a "cure point" (based on the rationale that alloHSCT is a potentially curative treatment) at 5 years and the application of a standardised mortality ratio (SMR) for background mortality. Clinical opinion was sought on when patients would be considered cured if they did not relapse after transplantation. **For discussion**	The company state that there is consensus among clinical experts that patients who survive alloHSCT and are disease-free for more than 5 years can be expected to be cured. Relapse and death rates after 5 years are rare. Most survival curves show a plateau at 5 years.	The approach chosen by the company to model mortality is plausible. The model is robust to changes in the fixed "cure point" except when the "cure point" is 1 year.

2.1 The technical team recognised that the following uncertainty would remain in the analyses and could not be resolved:

- The MC-FludT.14/L Trial II clinical trial only includes adults with AML and MDS who are at increased risk for standard conditioning therapies (that is, not eligible for standard myeloablative conditioning (MAC) busulfan- or total body irradiation (TBI)-based regimens) and does not provide evidence for people with other malignancies, people who might be eligible for MAC standard and for children.

2.2 The implementation of the ERG preferred assumptions resulted in treosulfan generating 0.78 more QALYs than busulfan at lower costs (-£17,689)[2]. Therefore, treosulfan dominated busulfan as in the company base-case. In the company base-case, treosulfan generated 0.89 more

[2] Following an update from the company, results presented in the committee meeting slides differ from the one in the technical report. The results presented in the slides include the company's price update.

QALYs than busulfan and saved £23,759[2]. In the ERG base-case, treosulfan provides both cost savings and QALY gains.

Table 1: ERG preferred deterministic base-case results (discounted) (Table 7.2 in ERG Report)

Technologies	Total costs (£)	Total LYGs	Total QALYs	Incr. costs(£)[3]	Incr. LYGs	Incr. QALYs	ICER versus baseline (£/QALY)
Treosulfan	£128,147	8.75	6.49	-£17,689	0.92	0.78	Treosulfan dominates
Busulfan	£145,836	7.84	5.71				

ERG = Evidence Review Group; ICER = incremental cost effectiveness ratio, Incr. = incremental, LYGs = life years gained, QALY = quality-adjusted life year

Table 2: ERG preferred probabilistic base-case results (discounted)

Technologies	Incr. costs (£)[3]	Incr. LYGs	Incr. QALYs	ICER versus baseline (£/QALY)
Treosulfan vs busulfan	-£15,857	0.97	0.70	Treosulfan dominates

ERG = Evidence Review Group; ICER = incremental cost effectiveness ratio, Incr. = incremental, LYGs = life years gained, QALY = quality-adjusted life year

2.3 Treosulfan is not likely to meet the end-of-life criteria

2.4 No equality issues were identified.

[3] Following an update from the company, incremental costs presented in the committee meeting slides differ from the one in the technical report. The results presented in the slides include the company's price update.

3. Key issues for consideration

Issue 1 - How does treosulfan differ from other alkylating agents used in conditioning regimens? What is the main benefit of treosulfan? - FOR DISCUSSION

Background/description of issue	• Treosulfan is an alkylating agent proposed as a well-tolerated alternative to other chemotherapy drugs (including other alkylating agents) in conditioning regimens for alloHSCT. • The primary endpoint of the MC-FludT.14/L Trial II is event-free survival (EFS). Events were defined as disease relapse, graft failure or death, whatever occurred first. Secondary endpoints included overall survival, non-relapse mortality, transplantation-related mortality and cumulative incidence of relapse. The efficacy results of the trial are reported in Appendix 1. • As stated by the company, a reduction in non-relapse mortality (NRM) is the reason for the overall survival benefit observed with treosulfan-based conditioning. However, this is included in the composite outcome of EFS as event free death and differences in relapse rates appear modest. The causes of the non-relapse deaths are unclear and could be part of non-relapse mortality (NRM) or transplantation-related mortality (TRM). • The modelling shows that treosulfan approximately halved the time in the relapse/progression health state compared to busulfan. However, it is not clear how this relates to clinical trial evidence as the relapse rates appear comparable between treosulfan and busulfan. It appears that the main difference between treosulfan and busulfan is observed in the death events of EFS.
Why this issue is important	The understanding of the main benefit of treosulfan and its mechanism of action is key for the appraisal of treosulfan.

Questions for engagement	1. What clinical advantages does treosulfan have over other alkylating agents used in conditioning regimens for alloHSCT? 2. What are the causes of death in the group of people who do not relapse (non-relapse mortality)- which in the company's submission is said to drive the OS benefit- are they non-cancer related deaths? 3. What is the difference between transplantation-related mortality (TRM) and non-relapse mortality (NRM)? 4. The main driver of the model is delay to relapse (rather than non-relapse death which gives the OS benefit), but the relapse rates appear similar between treosulfan and busulfan, is there any time to relapse data to clarify this point? 5. The modelling shows that treosulfan approximately halved the time in the relapse/progression health state compared to busulfan. How does this relate to the evidence from the trial?
Technical team preliminary judgement and rationale	The technical team welcome clinical opinion on the main benefit of treosulfan and clarification on whether it impacts the relapse of the disease or the transplantation-related mortality and its complications only.
Summary of comments	The clinical expert stated that: • Treosulfan is safer (leading to lower non-relapse mortality) and less toxic than busulfan and busulfan needs to be administered with antiepileptic prophylaxis. • The efficacy with treosulfan is similar The company stated that: • The main advantage of treosulfan is the lower toxicity compared with other regimens, especially the busulfan and fludarabine regimen which is one of the most frequently used conditioning regimens in many countries. There is better tolerability leading to reduced non-relapse mortality (NRM)/transplant-related mortality and improved EFS (+14.5%) and OS (+12.5%). Treosulfan combines the advantages of low toxicity regimen, is a soft MAC so it can be used in comorbid and older patients. • The main causes of non-relapse death include infections and graft-versus-host-disease (GvHD). Also, there were fewer deaths from infections (9.3% vs 14.1%) or GvHD (4.8% vs 7.4%) with treosulfan and fludarabine compared to busulfan and fludarabine. • Non-relapse mortality was defined as the probability of dying without previous occurrence of a relapse or progression of the underlying disease.

	• In contrast to non-relapse mortality (NRM), transplantation-related mortality (TRM) also includes patients who died from treatment while having a relapse/progression leading to TRM being slightly higher than NRM. • Differences between treosulfan and busulfan in the model are due to differences between OS and EFS curves, determining the proportion of patients in the relapse/progression/graft failure health state over time. EFS was significantly improved for treosulfan in the MC-FludT.14/L Trial II, which was driven by fewer deaths (13.1% vs 19.8%), fewer relapse/progression (22.8% vs 25.4%) and fewer graft failure (0.4% vs 3.2%). • The mean time to relapse/progression was longer in the treosulfan arm (235.6 days vs 212.5 days) as well as mean time to relapse/progression or graft failure (232.3 days vs 203.4 days). • The differences in OS and EFS between treatment arms define the distribution of patients in the different health states in the model. OS and EFS are extrapolated beyond the duration of the clinical trial. As the difference between OS and PFS is larger at the end of the trial for busulfan compared to busulfan, these differences observed in the trial are then preserved beyond the clinical trial duration and for a longer period of time for busulfan compared to treosulfan. In the ERG's preferred assumptions, the adjusted life table estimates are applied to the OS curves only, resulting in the OS becomes equal to EFS at approximately 8 years for treosulfan compared to 13.3 years for busulfan.
Technical team judgement after engagement	The technical team understand that the main benefit of treosulfan is reduced toxicity leading to lower non-relapse mortality and that there is limited effect on relapse rates. It is in unclear whether treosulfan is a myeloablative regimen at the dose administered in the trial and specified in the marketing authorisation.

Issue 2 - To what extent can the results of the trial be extrapolated to a broader population of patients with other malignancies? – FOR DISCUSSION

Background/description of issue	• The full marketing authorisation is "Treosulfan in combination with fludarabine is indicated as part of conditioning treatment prior to allogeneic haematopoietic stem cell transplantation (alloHSCT) in adult patients with malignant and non-malignant diseases, and in paediatric patients older than one month with malignant diseases." This appraisal focuses on malignant disease only and the population defined in the scope is "People with malignant disease that is in remission before allogeneic haematopoietic stem cell transplantation". Another appraisal is focusing on non-malignant disease (ID1540). • The company submitted evidence in a narrower population, which is the population in the main clinical trial MC-FludT.14/L Trial II. The trial population is people with AML or MDS who are at increased risk for standard conditioning therapies (that is, not eligible for standard myeloablative conditioning [MAC] busulfan- or total body irradiation [TBI]-based regimens). Patients ineligible for standard conditioning therapies were defined as those who were aged ≥50 years at transplant and/or who had haematopoietic cell transplantation co-morbidity index (HCT-CI) score≥2. • The company believes that this trial is the best evidence available for treosulfan and is applicable to a broader population of adults, children and young people with malignant disease that is in remission before alloHSCT. The company justified this assumption based on a position paper from six European clinicians (including 2 clinical experts from the UK). • The ERG is concerned that the population included in MC-FludT.14/L Trial II is not representative of the full population defined in the scope. Specifically, it did not include: o Adults with malignancies other than AML or MDS o People who can tolerate standard MAC o Children • The ERG considers that the evidence submitted can only support the assessment of treosulfan and fludarabine in adults with AML or MDS who are at increased risk for standard conditioning therapies. • In its submission, the company refers to another trial conducted in children with malignant diseases (MC-FludT.17/M) which could permit, in conjunction with MC-FludT.14/L Trial II data, to extend the use of treosulfan to children, by extrapolating efficacy and safety results. See Table 4.12: Summary of study

	methodology for the ongoing study in children (MC-FludT.17/M) on page 71 of the ERG report. • The CS stated that: "Overall, the reported safety and efficacy results of this Phase 2 alloHSCT trial demonstrated a positive benefit-risk for the treosulfan-based conditioning regimen used in the selected paediatric population and thus allowing extension of the use of treosulfan to this paediatric population." • The ERG does not consider that this ongoing study provides sufficient evidence to support the extension of the use of the treosulfan-based conditioning regimen to children with haematological malignancies, because it does not provide any indication of comparative effectiveness vs. other alternative conditioning regimens.
Why this issue is important	The evidence submitted does not cover the population defined in the scope and in the marketing authorisation (MA) for malignant diseases. The MA explicitly includes children. Although the company believes that the evidence of the MC-FludT.14/L Trial II is generalisable to the broader population of the scope, the population included in MC-FludT.14/L Trial II is not representative of the full population defined in the scope, specifically, there is a lack of evidence about the effectiveness of treosulfan and fludarabine conditioning regimens in adults with malignancies other than AML or MDS and in children.
Questions for engagement	1. Is it plausible to assume that the data from MC-FludT.14/L Trial II is generalisable to a broader population of people with malignant disease requiring conditioning treatment, specifically people with malignant diseases other than AML and MDS? 2. To what extent could the treatment-effect on endpoints such as time to relapse and death observed in the trial be relevant to the broader population of people with malignant disease (the Kaplan-Meier estimates of EFS are reported in Appendix 3)? 3. In clinical practice, is the efficacy of conditioning treatments such as treosulfan expected to be the same irrespective of the underlying malignant disease? 4. In clinical practice, is the efficacy of conditioning treatment expected to be the same in patients for whom MAC is not suitable and in patients who could have MAC? 5. In clinical practice, is the efficacy of conditioning treatment expected to be the same in adults as in children? 6. Is the MC-FludT.17/M trial sufficient to support the extension of the use of treosulfan to children? 7. Would additional cost-effectiveness analysis conducted with data from the MC-FludT.17/M (ongoing study in children) provide sufficient evidence to support the appraisal of treosulfan in children?

Technical team preliminary judgement and rationale	The technical team welcome clinical opinion on the clinical plausibility of the assumption that the benefits observed in MC-FludT.14/L Trial II are generalisable to a broader population of people with malignant diseases other than AML and MDS and in children.
Summary of comments	**The company** stated: • The trial results are generalisable to a broader population, because the main benefit of treosulfan is the reduced toxicity which is not impacted by the underlying disease. Evidence has been provided within the large EBMT registry study in patients with multiple myeloma[4]. • The efficacy of treosulfan-based conditioning has been shown in other malignancies as reported in the Haematologists' position paper submitted as Appendix L.7 written by six clinical transplant experts. • The company believes that the efficacy of treosulfan will be the same irrespective of the underlying malignant disease. The main purpose of conditioning is to remove the self-renewing capacity of the patient's own haematopoiesis and to suppress his immune system to allow the engraftment of donor stem cells. This is independent of the underlying disease. • Non-treosulfan-containing MAC regimens are not suited for older and/or comorbid patients because of their high toxicity. Treosulfan is the first MAC regimen that can be used in older and/or comorbid patients. It has also been studied in many trials that included patients without risk factors for standard MAC regimens. • European guidelines currently recommend treosulfan-based regimen as an option for children with MDS[5] and with primary immunodeficiencies[6]. • Outcomes in children with malignant disease undergoing alloHSCT are usually better than in adults, probably due to a lower frequency of comorbidities in children. The efficacy of treosulfan was demonstrated in the MC-FludT.17/M trial including 70 children with malignant diseases. A survival analysis updated was recently performed and showed a very low TRM (7.9% at 3 years) and low cumulative incidence of relapse/progression (22.4%) resulting in improved overall survival (86.7% at 3

[4] Gran, C. et al. Treosulfan Conditioning for Allogeneic Transplantation in Multiple Myeloma Improved Overall Survival in Upfront Hematopoietic Stem Cell Transplantation — a Large Retrospective Study By the Chronic Malignancies Working Party of the EBMT. Blood 132, 3464 (2018).
[5] EWOG-MDS. Guidelines for Hematopoietic Stem Cell Transplantation (HSCT) in Childhood MDS and JMML for Patients enrolled in EWOG-MDS Studies. Version 1.3. in EWOG-MDS Consensus Conference Freiburg 1–19 (2016)
[6] European Society for Blood and Marrow Transplantation & European Society for Immunodeficiencies. EBMT / ESID Guidelines for Haematopoietic Stem Cell Transplantation for Primary Immunodeficiencies. ESID EBMT HSCT Guidelines 2017 (2017)

	years) which is 20% better than in the MC-FludT.14/L Trial II in older and comorbid adults (66.8% at 3 years). • The company believes that the MC-FludT.14/L Trial II is sufficient to support the extension of the use of treosulfan in children. The efficacy of treosulfan in children has been confirmed in many trials. • The MC-FludT.17/M trial has now been completed and patients are currently in the follow-up period for survival status. The trial was a single-arm study and therefore only indirect treatment comparisons with other regimens could be made. The company decided not to perform an additional cost-effectiveness analysis as the evidence is less robust than in adults and the same regimen busulfan and fludarabine is used as a comparator.
Technical team judgement after engagement	The technical team considers that if the main benefit of treosulfan is on the toxicity and not on disease relapse, then the assumption that benefits are generalisable to a broader population of adults with other malignant diseases is plausible. Regarding children the company refers to the MC-FludT.17/M trial, the EBMT registry data and a meta-analysis to support the efficacy of treosulfan in children. However, no cost-effectiveness analysis is available.

Issue 3 - The submission focuses on a low-intensity regimen – FOR DISCUSSION

Background/description of issue

- The evidence for the clinical effectiveness of treosulfan and fludarabine is based on one comparator (low dose busulfan with fludarabine, which is the comparator of the MC-FludT.14/L Trial II trial) therefore there is limited evidence on the comparative effectiveness of treosulfan and fludarabine versus other conditioning regimens particularly for standard high-intensity MAC regimens.
- The table below summarise the comparators listed in the NICE scope, the comparators listed in the company's decision problem and the comparators effectively included in the cost-effectiveness analysis.

Comparators in NICE scope	Included in company's decision problem	Included in the cost-effectiveness analysis
Standard high-intensity (myeloablative) conditioning regimens		
cyclophosphamide and total body irradiation	Yes	No
cyclophosphamide and busulfan	Yes	No
cyclophosphamide and thiotepa	No	No
high dose busulfan with fludarabine with or without thiotepa	No	No
Reduced intensity conditioning regimens		
low dose busulfan with fludarabine	**Yes**	**Yes**
melphalan plus fludarabine	No	No

- The company explained that the submission focused on the reduced intensity conditioning because this is where they have direct comparative evidence from the MC-FludT.14/L Trial II trial, and also where there is an unmet need for a large proportion of patients. Moreover, the company explained that the primary effect of thiotepa is to provide additional immunosuppression thereby preventing graft rejection but does not enhance stem cell depletion. Therefore, thiotepa is always administered in combination with a stem cell toxic treatment such as treosulfan, busulfan or melphalan.
- As a result, the ERG considers that the evidence submitted is sufficient to support an assessment of the cost effectiveness of treosulfan and fludarabine versus RIC busulfan and fludarabine in adults with AML

	or MDS, who are at increased risk for standard conditioning therapies (as explained in Issue 2) but is not sufficient to support an assessment versus any of the other comparators defined in the scope.
Why this issue is important	The comparators are crucial to the consideration of the clinical and cost-effectiveness evidence. As stated in the methods guide for technology appraisal, the most appropriate comparators should reflect the established NHS practice in England, the natural history of the condition without a suitable treatment, and existing NICE guidance. Furthermore, the nature of the restricted population in MC-FludT.14/L Trial II (adults who were at increased risk and therefore not eligible for standard MAC conditioning regimens), and hence the choice of comparator (RIC busulfan and fludarabine) means that the relative effectiveness and cost effectiveness of treosulfan and fludarabine has not been evaluated against the range of comparator regimens that would be relevant for the full population defined in the scope.
Questions for engagement	1. Is the company's choice of comparator appropriate? 2. Is reduced intensity conditioning (RIC) the area with the largest unmet need? 3. What are the conditioning treatments used in clinical practice in people at high risk for standard conditioning therapy (that is, not eligible to standard high-intensity MAC)? 4. Are the comparator regimens all similarly effective? 5. For people at high risk for standard conditioning therapy who are eligible for reduced intensity regimens, is low dose busulfan with fludarabine as effective as melphalan with fludarabine? Is there any quantitative evidence (such as, clinical trials, registries) to support it? If not, can the clinical experts provide an estimate for how different it is? 6. For people at lower risk who are eligible for standard regimens, what is the difference in time to event between standard high-intensity regimens and low dose busulfan? Is there any quantitative evidence available? If not, can the clinical experts provide an estimate for how different it is? 7. Do the regimens all have similar costs (e.g. acquisition cost, resource use for administration etc.)?
Technical team preliminary judgement and rationale	For patients at increased risk for standard conditioning therapy, the company's approach and choice of comparator is reasonable as the comparators of interest would be low dose busulfan and fludarabine and melphalan/fludarabine. However, the reason for excluding melphalan/fludarabine (another reduced intensity conditioning regimen) is not clearly justified. There is a lack of evidence about the comparative effectiveness of treosulfan and fludarabine conditioning regimens in people who can tolerate standard MAC regimens (see Issue 5).

Summary of comments	**The company** stated: • The comparator's choice is appropriate because busulfan and fludarabine (RIC) is the most frequently used conditioning regimen in older and/or comorbid patients. • The area of largest unmet need is not reduced intensity conditioning, but it is the reduction of toxicity with MAC regimen, that is, developing reduced-toxicity standard intensity conditioning regimen that can be used in older and/or comorbid patient. • Most adult patients with haematological malignancies are older than 60 years and often suffer from comorbidities. AlloHSCT is usually the only chance of cure, however the RIC regimen is associated with a higher risk of relapse. Treosulfan and fludarabine is a reduced-toxicity MAC regimen aiming to address this unmet need of reduced toxicity without impacting efficacy. • There is no consensus on conditioning regimen for patients not eligible for standard high-intensity MAC regimens. The RIC regimens the most frequently used currently are low dose busulfan plus fludarabine and melphalan plus fludarabine. • Furthermore, results from the CIBMTR analysis showed nearly equivalent EFS and OS rates with various non-treosulfan regimens (low or high dose busulfan and fludarabine, busulfan and cyclophosphamide, fludarabine and melphalan). • People at lower risk (age<50 years, no significant comorbidities) are preferably treated with standard MAC regimen, as low intensity regimens are associated with a high risk or relapse. There is insufficient evidence to generalise about the comparative benefit of MAC versus RIC in younger patients without comorbidities. These patients tolerate MAC regimen and clinicians are unlikely to choose a RIC regimen which would increase the risk of relapse. • The total costs for all regimens are not all similar, varying from £901.89 to £8,956.67 (excluding wastage, see Appendix 4 for full table). Individual conditioning regimens are unlikely to be significant drivers of overall costs given the short treatment period. In the economic analysis, treatment costs were less than 3% of the overall costs even when assuming a 100% wastage. • In terms of resource use for treatment administration, the total number of administrations varies from 6 to 11 for all treatments, suggesting that the differences in resource use requirements for treatment administration may be relatively small and are unlikely to represent a large proportion of the overall costs compared to costs generated through differences in clinical efficacy.

Technical team judgement after engagement	Busulfan is the most common regimen used in the UK and a good proxy for conditioning regimens used in the UK according to clinical opinion. Treosulfan, at a dose of 10 mg/m^2, in combination with fludarabine, has only been compared with a reduced intensity regimen using low dose busulfan. There is no evidence which to the comparative efficacy of this treosulfan regimen versus MAC regimens (including higher doses of busulfan) which are used for people who can tolerate these high intensity regimens.

Issue 4 - Generalisability of the trial results to the predicted population for whom it might be used in the NHS (that is, people who would otherwise get reduced intensity busulfan) - AGREED

Background/description of issue	• There were no UK patients in the MC-FludT.14/L Trial II trial which was the basis of the clinical effectiveness evidence for treosulfan and fludarabine • Most patients included in the trial were from Germany. The mean age was 59.6 years and mean weight was 80.2kg. (see appendix 2 for the trial patient characteristics) • In the MC-FludT.14/L Trial, included patients were defined as being at increased risk for standard conditioning therapy (that is, not eligible for standard high-intensity MAC). • The clinical trial definition of increased risk (as defined by the trial's inclusion criteria) was: patients aged 50 years and older and/or with a hematopoietic cell transplantation co-morbidity index (HCT-CI) score>2. It is not clear to the ERG how this definition would be similar to the definition of patients at increased risk generally applied in UK clinical practice. The ERG believes that a consistent definition should be established in order to inform any recommendation for this population. • According to the latest British Society of Blood and Marrow transplantation (BSBMT) registry (2017), the most common indications for an alloHSCT in the UK are AML (36%), acute lymphoblastic leukaemia (ALL, 16%) and myelodysplastic syndrome (MDS) or myelodysplastic/ myeloproliferative neoplasms (MDS/MPN; 13%). • The company commissioned the analysis of the US Centre for International Blood and Marrow Transplant Research (CIBMTR) to compare it to the data from the MC-FludT.14/L Trial II. Registry patients included the analysis were aged ≥50 years or aged 18 to 70 years with HCT-CI score >2, which is in line with the trial's definition of increased risk patients not eligible for high-intensity MAC

	conditioning regimen. However, the conditioning regimen received by patients in this registry included some high-intensity MAC such as busulfan and cyclophosphamide, busulfan and fludarabine. As a result, the ERG is unclear on whether the MC-FludT.14/L Trial II used additional criteria to define patients at increased risk for standard therapies. • Also, the company believe that the general practice of alloHSCT in England and Wales is not different from other major European countries including France, Germany, Hungary, Italy and Poland. This is demonstrated in the most recent European Society for Blood and Marrow Transplantation (EBMT) report showing 4,316 alloHSCT performed by 52 UK teams in 2016. These transplant sites in the UK are part of the EBMT and as a result, treat their patients according to the EBMT guidelines. • The ERG agrees that alloHSCT practice is likely to be broadly similar in England and Wales to that in other European countries. However, the ERG is unclear as to how similar the definition of patients at increased risk for standard conditioning therapies is to how they are defined in UK clinical practice.
Why this issue is important	The clinical effectiveness evidence should be generalisable to the UK as the technology recommendation will be for UK practice.
Questions for engagement	1. Is it reasonable to assume that alloHSCT practice is broadly similar in England and Wales to that in other European countries? 2. Is it plausible that as UK transplant sites are members of the EBMT, they also treat their patients according to the EBMT Guidelines? 3. Is the population of MC-FludT.14/L Trial II broadly representative of UK clinical practice in terms of age and weight? 4. In UK clinical practice, how are people at increased risk (not eligible to standard MAC) defined? Is this similar to how they were defined in the clinical trial? 5. Is it plausible to assume that the time to relapse and time to death observed in the trial for those who received low dose busulfan with fludarabine is generalisable to patients in the UK who receive this regimen (event-free survival at 24 months was 51.2% in the busulfan arm, EFS curves are reported in Appendix 3)? 6. Which patient's characteristics would be expected to be treatment-effect modifiers?
Technical team preliminary judgement and rationale	The technical team would welcome clinical opinion on the comparability of clinical practice between the UK and other European countries as well as the comparability between clinical trial population and the UK indicated population.

Summary of comments	The company stated that: • It is reasonable to assume that alloHSCT practice is similar in England in Wales as with other European countries, because 52 UK transplant centres are members of the EBMT and work according to EBMT guidelines. • More than 20 UK experts were involved in the development and writing of the 2019 EBMT Handbook for Haematopoietic Stem Cell Transplantation and Cellular Therapies[7]. • The MC-FludT.14/L study reflects the routine clinical practice in the UK, and it is plausible to assume that the time to relapse and death observed in the trial is generalisable to patients in the UK who receive this regimen. However, there are only limited published EFS and OS data for busulfan and fludarabine RIC from the UK. • A prospective study in 75 UK patients with MDS receiving busulfan and fludarabine RIC showed a 3-year OS of 43% and disease-free survival of 41%, which is about 10% lower than what was observed in the MC-FludT.14/L trial (56.3% and 49.7% respectively) (Lim 2006[8]) • The most important treatment-effect modifiers are age, disease status at transplant, risk group, donor type (MRD or MUD), pre-treatment and comorbidity index.
Technical team judgement after engagement	The technical team agree that it is reasonable to assume that alloHSCT practice is similar in England in Wales as with other European countries.

[7] European Society for Blood and Marrow Transplantation (EBMT). The EBMT Handbook Hematopoietic Stem Cell Transplantation and Cellular Therapies. (2019).
[8] Lim, Z. Y. et al. Outcomes of alemtuzumab-based reduced intensity conditioning stem cell transplantation using unrelated donors for myelodysplastic syndromes. Br. J. Haematol. 135, 201–9 (2006).

Issue 5 - Limited evidence of comparative effectiveness of treosulfan versus other conditioning regimens particularly myeloablative conditioning (MAC) regimens – FOR DISCUSSION

Background/description of issue	• As stated in Issue 2, the only direct comparative data available for treosulfan and fludarabine is the comparison to busulfan and fludarabine in the MC-FludT.14/L Trial II. • The company explored 2 approaches to obtain comparative evidence; registry analyses and indirect treatment comparison: **Registry analyses** • The company commissioned the analysis of two registry studies, the European Society for Blood and Marrow Transplantation (EBMT) and Centre for International Blood and Marrow Transplant Research (CIBMTR), in order to compare patients from MC-FludT.14/L Trial II who received treosulfan and fludarabine to registry patients who received other conditioning regimens. These analyses only included registry patients who matched the inclusion criteria in MC-FludT.14/L Trial II. However, these analyses were not used in the cost effectiveness modelling. • The ERG noted that there were limitations with these analysis such as the CIBMTR did not use an appropriate statistical method as there was no matching and only variables chosen by statistical selection were included in the final model, it was not based on clinical knowledge and differences between the treatment groups. Furthermore, conclusions from both registry analyses are limited by the fact that these are comparisons of two groups from different studies (registry and prospective randomised controlled trials [RCT]) which have been matched using statistical methods and not randomisation. The analyses can only include variables which have been measured in both studies, other important variables may differ between groups but cannot be included. The studies may also have important differences in patient inclusion criteria and data collection methods. These analyses were not used in the cost effectiveness modelling. **Indirect treatment comparison (ITC)** • The company conducted a feasibility assessment for the completion of indirect and mixed treatment comparisons to provide comparative data of treosulfan and fludarabine compared to other comparators included in the scope. This was conducted on 4 RCT reporting regimens of interest and outcomes of interest (Overall survival (OS), Event-free survival (EFS), RR (relapse rate) and Graft-versus-host

	disease (GvHD). • Following feasibility assessment, the company concluded that some indirect treatment comparisons were possible for treosulfan and fludarabine versus busulfan/cyclophosphamide and busulfan/fludarabine (MAC) at 2 years for OS, RR and the incidence of GvHD. • However, the company considered that these outcomes were unlikely to be reliable enough to provide relevant comparative data, as there were differences in trial and patient's population which may affect the validity of the ITC (e.g. the proportion of patients receiving unrelated alloHSCT). • At clarification, the ERG requested these analyses to be conducted. However, the company did not conduct the analyses and believe that some of the endpoints are not sufficiently informative for the submission. Moreover, as the economic model is based on both EFS and OS but the ITC cannot provide evidence for EFS, the ITC was not performed. • As a result, the only comparative evidence included in the modelling is the direct comparison of treosulfan and fludarabine and busulfan and fludarabine from the MC-FludT.14/L Trial II. • The ERG considers that any analyses that may provide comparative evidence for treosulfan and fludarabine versus other regimens included in the NICE scope should be performed, wherever possible. The ERG considers that an analysis of OS could be possible but acknowledge the differences between trial populations.
Why this issue is important	There is a lack of evidence about the effectiveness of treosulfan and fludarabine conditioning regimens in people who can tolerate standard MAC regimens. The included RCT, MC-FludT.14/L Trial II, only provides comparative efficacy data for treosulfan in combination with fludarabine versus one alternative conditioning therapy, a RIC regimen of busulfan in combination with fludarabine. The company's submission did not include any indirect comparisons. An indirect treatment comparison of treosulfan and fludarabine vs other comparators of the scope could enable their inclusion in the cost-effectiveness analyses and provide evidence to support the appraisal of treosulfan and fludarabine versus more comparators. The registry analysis (if improved, noting the limitations stated by the ERG in section 4.2.9 in the ERG report) could also provide additional evidence to support the appraisal of treosulfan and fludarabine versus more comparators.

Questions for engagement	1. Would clinicians consider that the evidence from MC-FludT.14/L Trial II is good enough for treosulfan and fludarabine to be used in patients who can tolerate MAC? Or would patients who can tolerate MAC receive MAC regimen anyway? 2. Should treosulfan and fludarabine be compared with MAC regimens? 3. Could the EBMT and CIBMTR analyses be improved (for example by adjusting for age, matching and incorporating clinical opinion to identify the clinically significant variables to adjust for) while noting the limitations stated by the ERG, be incorporated into the CE model as a scenario to provide more robust/improved evidence of comparative effectiveness of treosulfan versus other conditioning regimens? 4. Do the differences between the trials included in the ITC justify the company's decision not to conduct an ITC? 5. Are these patients' characteristics of the trial expected to be treatment-effect modifiers for example, number of patients receiving matched unrelated alloHSCT and age?
Technical team preliminary judgement and rationale	The technical team believe that the evidence submitted does not support the use of treosulfan in people who can tolerate MAC regimens.
Summary of comments	The company stated that: • Clinicians would consider the evidence is sufficient to use treosulfan and fludarabine in MAC eligible patients. The company believe that treosulfan and fludarabine is a MAC regimen with a reduced toxicity compared to other MAC regimens. • For AML and MDS patients, the efficacy of treosulfan and fludarabine has been confirmed in the MC-FludT.7/AML and MC-FludT.8/MDS trials and data from about 1,000 patients documented in the EBMT registry[9,10]. • Published retrospective studies from the EBMT registry comparing treosulfan and fludarabine to other MAC regimens demonstrated that treosulfan-based regimen is at least as good as other MAC regimens.

[9] Nagler, A. et al. Long-term outcome after a treosulfan-based conditioning regimen for patients with acute myeloid leukemia: A report from the Acute Leukemia Working Party of the European Society for Blood and Marrow Transplantation. Cancer 123, 2671–2679 (2017).
[10] Shimoni, A. et al. Fludarabine and treosulfan conditioning is associated with a more favorable outcome after allogeneic stem cell transplantation in myelodysplastic syndrome. A survey on behalf of the Chronic Malignancies Working Party of EBMT.No Title. in 41st Annual Meeting of the EBMT (2015).

	• For example, Shimoni et al.[10] performed a retrospective analysis of all alloHSCTs for MDS performed between 2000 and 2011 in which they identified 480 patients who had treosulfan and fludarabine and compared their outcomes to patients with various MAC and RIC. The authors concluded that treosulfan and fludarabine is associated with similar low relapse rates as other MAC regimen and similar low non-relapse mortality as RIC regimens, resulting in improved outcome over both RIC and MAC. • It is difficult to compare studies published with other conditioning regimen to the results of the MC-FludT.14/L trial results because patient's characteristics were largely different. • Patient's characteristics in the trial are expected to be treatment-effect modifiers. Please note that the benefits of treosulfan-based conditioning have been shown in matched unrelated donors (MUD) as well as matched related donors (MRD) and in patients of different age groups
Technical team judgement after engagement	The technical team believe that the evidence submitted does not support the use of treosulfan in people who can tolerate MAC regimens.

Issue 6 - Mortality modelling – FOR DISCUSSION

Background/description of issue	• In order to model mortality, the company used a "cure point" based on the rationale that alloHSCT is a potentially curative treatment and applied a standardised mortality ratio (SMR) for background mortality. • Prior to the "cure point", patients transition between health states based on the parametric EFS and OS curves. After the "cure point", transition between health states are determined by the background mortality only. • In the company's base-case, a fixed "cure point" of 5 years is assumed, which is based on 2 clinical experts' opinion. At clarification, the company explained that patients who survive alloHSCT for at least 5 years are considered cured, meaning that relapses or transplant-related deaths are very rare after 5 years. • The impact on the results of choosing different "cure points" was tested by the ERG. The results were similar to the base-case except when the "cure point" was assumed to be one year, where busulfan was not dominated by treosulfan and resulted in treosulfan generating fewer QALYs than busulfan but also

	at a lower cost than busulfan. This result therefore generated an ICER of £47,910/QALY for busulfan compared to treosulfan[11]. Assuming a "cure point" of 3 years provided the maximum incremental QALYs and minimum savings in costs. For "cure point" higher than 5 years, the results were approximately the same as in the base-case. • Once the patients reach the "cure point", they progress according to HSCT-specific mortality rates. These mortality rates are calculated by using UK life tables and applying a SMR for alloHSCT. The ratio selected for the base-case analysis was the HR used in NICE TA552 which was considered a plausible estimate by the ERG in TA522 (HR=2.30) and by the clinical experts consulted by the company. The impact of choosing another SMR was not explored by the company and the ERG conducted additional scenario analyses. • The approach used in the base-case is to use parametric curves up to the fixed "cure point" and then to switch to background mortality that is HSCT-specific (using SMR-adjusted life tables). In the model, the cure point refers to patients that are functionally cured. Therefore, some patients can relapse after the "cure point". This implies that the underlying assumption is that when patients have survived five years after alloHSCT, most of the mortality risk is not attributed to a relapse of AML or MDS but to other causes such as long-term complications associated with alloHSCT itself.
Why this issue is important	It is important to understand how mortality is modelled and whether the approach is clinically plausible.
Questions for engagement	1. If patients did not relapse after transplantation, would they be considered cured at 1 year, 2 years, 5 years?
Technical team preliminary judgement and rationale	The approach chosen by the company to model mortality seems plausible. The model is robust to changes in the fixed "cure point" except when the "cure point" is 1 year. The technical team welcome clinical opinion on the clinical plausibility of the five-year fixed cure point.
Summary of comments	The company stated that: • There is consensus among clinical experts that patients who survive alloHSCT and are disease-free for more than 5 years can be expected to be cured. Relapse and death rates after 5 years are rare. • Most survival curves show a plateau at 5 years, In the MC-FludT.14/L Trial, the plateau was reached after 40 months. The 5-year functional cure point was suggested by the two clinical experts

[11] Following an update from the company, results presented in the committee meeting slides differ from the one in the technical report. The results presented in the slides include the company's price update.

	interviewed for the economic model validation.
Technical team judgement after engagement	The approach chosen by the company to model mortality is plausible. The model is robust to changes in the fixed "cure point" except when the "cure point" is 1 year.

4. Issues for information

Tables 1 to 3 are provided to stakeholders for information only and not included in the technical report comments table provided.

Table 1: ERG's preferred assumptions and impact on the cost-effectiveness estimate

The implementation of the ERG preferred assumptions resulted in treosulfan generating 0.78 more QALYs than busulfan at lower costs (-£17,689).

Alteration	Technical team rationale	Incremental costs[12]	Incremental QALYs	ICER
Company base case	–	-£23,759	0.89	Treosulfan dominates
Company's base case after clarification	-	-£23,668	0.89	Treosulfan dominates
ERG correction of OS and EFS implementation	Technical team agrees with ERG's amendments (See Table 3)	-£14,492	0.84	Treosulfan dominates
ERG correction of rescaling factor (year to day)	Technical team agrees with ERG's amendments (See Table 3)	-£14,490	0.84	Treosulfan dominates
Using NMCM Weibull to model OS	See Issue 5	-£17,641	0.78	Treosulfan dominates

[12] Following an update from the company, incremental costs presented in the committee meeting slides differ from the one in the technical report. The results presented in the slides include the company's price update.

Alteration	Technical team rationale	Incremental costs[12]	Incremental QALYs	ICER
Using most recent life tables	Technical team agrees with ERG's amendments (See Table 3)	-£17,689	0.78	Treosulfan dominates
Cumulative impact of the ERG's preferred assumptions on the cost-effectiveness estimate (deterministic)	**–**	**-£17,689**	**0.78**	**Treosulfan dominates**
Cumulative impact of the ERG's preferred assumptions on the cost-effectiveness estimate (probabilistic)		**-£15,857**	**0.70**	**Treosulfan dominates**

Table 2: Outstanding uncertainties in the evidence base

Area of uncertainty	Why this issue is important	Likely impact on the cost-effectiveness estimate
The evidence submitted is narrower than the full scope population	The MC-FludT.14/L Trial II trial is based on adult patients with AML and MDS patients only, who are at increased risk for standard conditioning therapies (that is, not eligible for standard myeloablative conditioning (MAC) busulfan- or total body irradiation (TBI)-based regimens) and is not in line with the full scope population which is for people with malignant disease that is in remission prior to allogeneic hematopoietic stem cell transplant (alloHSCT).	Unknown
The evidence submitted refers to the comparison of treosulfan to one regimen only.	The evidence for the clinical effectiveness of treosulfan and fludarabine is based on one comparator (low dose busulfan with fludarabine, which is the comparator of the MC-FludT.14/L Trial II trial) therefore there is limited evidence on the comparative effectiveness of Treosulfan/Fludarabine versus other conditioning regimens particularly for standard	Unknown

Area of uncertainty	Why this issue is important	Likely impact on the cost-effectiveness estimate
	high-intensity MAC regimens. It is unclear whether the evidence is generalisable to the comparisons with other regimens.	

Table 3: Other issues for information

Issue	Comments
Event-free and overall survival extrapolation	• The company chose a non-mixture cure model (NMCM) lognormal for all analyses and all treatment arms • The ERG highlighted that the final model selection appeared to be based solely on AIC/BIC which can be misleading • The ERG conducted its own goodness-of-fit assessment of the models and concluded that the company's choice of lognormal NMCM for EFS was reasonable. However, it reported that this model seems to overestimate treosulfan EFS. As a result, the ERG conducted a scenario analysis using a gamma distribution for EFS for treosulfan but acknowledged that choosing different parametric distribution for each treatment arm is not recommended by the NICE DSU 14. • For OS, the ERG prefers to use the NMCM Weibull distribution instead of the company's choice of NMCM lognormal. This is because the NMCM lognormal seem to underestimate OS for busulfan despite having a lower AIC than the Weibull. The ERG conducted scenario analyses using the mixture-cure model (MCM) Weibull and MCM/NMCM lognormal as they also provided plausible estimates. • Using the ERG's preferred model and scenarios lead to results similar to the company's base-case in which busulfan is dominated by treosulfan. However, it resulted in less cost savings and smaller incremental QALYs

Issue	Comments
Subgroup analyses	• The company provided subgroup analyses for AML and MDS separately as a scenario analysis. The ERG questioned the rationale for including subgroup analyses as it is not included in the scope and more importantly because the company stated that the underlying disease is not the main determinant of treosulfan efficacy. • The company explained that these analyses were provided for validation purposes in order to check that the results within each population were consistent with the base-case results. • When performing these analyses, the company used the same distribution as for the pooled population (NMCM lognormal). The ERG highlighted that does not seems consistent with the methodology used in the base-case (lowest AIC/BIC and visual inspection) and that the NMCM lognormal does not fit the data for the MDS subgroup • The ERG believes that subgroup analyses should be performed with subgroup-specific input parameters including patient's characteristics and model selection should be performed for each subgroup. • The ERG conducted its own analyses for AML and MDS separately • For AML, the ERG performed analyses using the MCM lognormal distribution for OS and the NMCM lognormal distribution for EFS. Treosulfan dominated busulfan but both cost savings and incremental QALYs for treosulfan were lower than in the company's analysis. • For MDS, the ERG performed analyses using the NMCM Weibull distribution for OS and the MCM lognormal distribution for EFS. The results also showed that treosulfan was dominant, but cost savings were higher and incremental QALYs were lower than in the company's analysis.
OS and EFS probability calculations	Overall and event-free survival probabilities were incorrectly calculated using mortality rates instead of transition probabilities. The ERG amended the calculations in the model, but it had a minor effect on the results because the mortality rates (R) are small and the transition probabilities (calculated as 1-EXP(-R)) are similar to the mortality rates

Issue	Comments
Yearly values rescaled to daily values	Yearly values in the model where re-scaled to daily values using a factor 1/364, which was used in life tables, the calculation of health state costs and health state utilities. The ERG understands the company's assumption a year is assumed to have 364 days in the economic model with a cycle of 28 days. However, this assumption does not necessarily hold for the yearly values obtained from life tables, the calculation of health state costs or health state utilities. Therefore, the ERG prefers a factor 1/365.25 and then multiply this by 28 to adjust for the cycle length. This was a minor error and had no impact on the model results.
Most recent UK life tables used	The company did not used the most recent UK life tables, the ERG corrected this and amended calculation using the 2015-2017 UK life tables.
Calculation of qx in life tables	The calculation of qx (that is, the mortality rate between age x and x +1) in the model did not correspond with the formula used in the in the UK life tables. The ERG amended the calculations.
End-of-life criteria	Treosulfan is not likely to meet the end-of-life criteria which are the following: The treatment provides an extension of more than an average of three months compared to current NHS treatment and; The treatment is indicated for patients with short life expectancy, normally a mean life expectancy of less than 24 months The company have not made a case for end of life.

Authors

Jane Adam

Appraisal committee chair

Caroline Bregman

Technical lead

Rufaro Kausi

Technical adviser

Janet Robertson

Associate director

With input from the lead team:

Rita Faria

Lead team member

Min Ven Teo

Lead team member

Richard Ballerand

Lead team member

Appendix 1

Efficacy results of MC-FludT.14/L Trial II (from ERG report, Table 4.7)

	Treosulfan (10 g/m²/day) + Fludarabine (30 g/m²/day)	Busulfan (3.2 mg/kg/day) + Fludarabine (30 g/m²/day)
Number randomised	280	290
Number analysed (FAS: patients who received conditioning treatment and HSCT)	268	283
Median follow-up[a], months (range of those surviving)	29.7 (3.0 to 52.1)	29.4 (3.0 to 54.3)
Primary outcome – Event-free survival (EFS) within 24 months after alloHSCT		
Patients with event	97 (36.2%)	137 (48.4%)
Death[b]	35 (13.1%)	56 (19.8%)
Relapse/progression[b]	61 (22.8%)	72 (25.4%)
Primary graft failure[b]	1 (0.4%)	1 (0.4%)
Secondary graft failure[b]	0 (0.0%)	8 (2.8%)
Event-free survival at 12 months[c] [%] (95% CI)	70.0 (64.1 to 75.1)	60.8 (54.9 to 66.3)
Event-free survival at 24 months[c] [%] (95% CI)	65.7 (59.5 to 71.2)	51.2 (45.0 to 57.0)
Event-free survival at 36 months[c] [%] (95% CI)	[redacted]	[redacted]
Hazard ratio [HR][d] (95% CI)	0.64 (0.49 to 0.84)	
Secondary outcome – Overall survival (OS) within 24 months after alloHSCT		
Patients with event	81 (30.2%)	112 (39.6%)
Overall survival at 12 months[c] [%] (95% CI)	77.8 (72.3 to 82.3)	71.8 (66.1 to 76.7)
Overall survival at 24 months[c] [%] (95% CI)	72.7 (66.8 to 77.8)	60.2 (54.0 to 65.8)

	Treosulfan (10 g/m²/day) + Fludarabine (30 g/m²/day)	Busulfan (3.2 mg/kg/day) + Fludarabine (30 g/m²/day)
Overall survival at 36 months[c] [%] (95% CI)	[redacted]	[redacted]
HR[d] (95% CI)	0.64 (0.48 to 0.87)	
Secondary outcome – Cumulative incidence of relapse/progression 24 months after alloHSCT		
Patients with event	61 (22.8%)	72 (25.4%)
Patients without event (censored) or with competing event	207 (77.2%)	211 (74.6%)
Censored	171 (63.8%)	146 (51.6%)
Death[b]	35 (13.1%)	56 (19.8%)
Primary graft failure[b]	1 (0.4%)	1 (0.4%)
Secondary graft failure[b]	0 (0.0%)	8 (2.8%)
Cumulative incidence at 12 months [%] (95% CI)	19.1 (14.4 to 23.8)	21.7 (16.9 to 26.5)
Cumulative incidence at 24 months [%] (95% CI)	22.0 (16.9 to 27.1)	25.2 (20.0 to 30.3)
Cumulative incidence at 36 months [%] (95% CI)	[redacted]	[redacted]
HR[e] (95% CI)	0.82 (0.59 to 1.16)	
Secondary outcome - engraftment		
Primary graft failure[f]	1/268 (0.4%)	1/283 (0.4%)
Secondary graft failure[f]	0/263 (0.0%)	8/279 (2.9%)
Secondary outcome (not specified in scope) – Cumulative incidence of non-relapse mortality (NRM) 24 months after alloHSCT		
Patients with event	35 (13.1%)	56 (19.8%)
Patients without event (censored) or with competing event	233 (86.9%)	227 (80.2%)
Censored	171 (63.8%)	146 (51.6%)
Relapse/Progression[b]	61 (22.8%)	72 (25.4%)

	Treosulfan (10 g/m²/day) + Fludarabine (30 g/m²/day)	Busulfan (3.2 mg/kg/day) + Fludarabine (30 g/m²/day)
Primary Graft Failure[b]	1 (0.4%)	1 (0.4%)
Secondary Graft Failure[b]	0 (0.0%)	8 (2.8%)
Cumulative incidence at 12 months [%] (95% CI)	10.5 (6.8 to 14.2)	14.3 (10.2 to 18.4)
Cumulative incidence at 24 months [%] (95% CI)	12.0 (8.0 to 15.9)	20.4 (15.5 to 25.2)
Cumulative incidence at 36 months [%] (95% CI)	[redacted]	[redacted]
HR[g] (95% CI)	0.63 (0.41 to 0.97)	
Secondary outcome (not specified in scope) – Transplantation-related mortality (TRM)		
Patients with event	33 (12.3%)	58 (20.5%)
Patients without event	235 (87.7%)	225 (79.5%)
Transplantation-related mortality at 12 months[c] [%] (95% CI)	11.7 (8.3 to 16.3)	16.2 (12.2 to 21.3)
Transplantation-related mortality at 24 months[c] [%] (95% CI)	12.8 (9.2 to 17.7)	24.1 (19.1 to 30.2)
HR[d] (95% CI)	0.52 (0.34 to 0.82)	

Based on Tables 18, 19, 20, 21 and 23 of the CS[1]

[a] Based on reverse Kaplan-Meier estimates for overall survival; [b] Only if this event occurred first; [c] Based on Kaplan-Meier estimates; [d] Adjusted for donor type as factor, and risk group and centre as strata using Cox regression model; [e] Adjusted for donor type as factor and risk group as stratum using Fine and Gray model; [f] Rate of primary/secondary graft failure calculated as number of patients with graft failure by the number of patients at risk; [g] Adjusted for donor type as factor

alloHSCT = allogeneic haematopoietic stem cell transplantation; CI = confidence interval; CS = company submission; EFS = event-free survival; FAS = full analysis set; HR = hazard ratio; HSCT = Haematopoietic stem cell transplantation; NRM = non-relapse mortality; OS = Overall survival; TRM = transplantation-related mortality

Appendix 2

Patients characteristics from MC-FludT.14/L Trial II – pooled AML and MDS patients (from company's submission, Table 36)

Variable	Busulfan	Treosulfan	Total
N	283	268	551
N (male)	173	162	335
N (female)	110	106	216
Sex (male)	61.13%	60.45%	60.80%
Sex (female)	38.87%	39.55%	39.20%
Age (mean)	59.9	59.3	59.6
Age (SD)	6.0	6.5	6.3
Weight (mean), kg	79.4	80.9	80.2
Weight (SD), kg	17.7	16.7	17.3
n (MRD)	68	62	130
n (MUD)	215	206	421
% (MRD)	24.03%	23.13%	23.59%
% (MUD)	75.97%	76.87%	76.41%
BSA (mean), m^2	1.921	1.942	1.931
BSA (SD), m^2	0.241	0.227	0.235
RBC transfusion dependency, n (No)	219	216	435
RBC transfusion dependency, n (Yes)	64	52	116
RBC transfusion dependency, % (No)	77.39%	80.60%	78.95%
RBC transfusion dependency, % (Yes)	22.61%	19.40%	21.05%

Appendix 3

Figure 1 Kaplan Meier estimates of EFS - MC-FludT.14/L Trial II

Source: Company's submission Figure 4

Appendix 4 Update of results of trial MC-FludT.17/M (Executed: 09-Jan-2019)

Summary results of transplantation-related mortality (Full Analysis Set)	Treosulfan (N = 70)
Subjects with event	[redacted]
Subjects without event	[redacted]
Transplantation-related mortality at 100 days[a] [%] 90% CI	[redacted]
Transplantation-related mortality at 12 months[a] [%] 90% CI	[redacted]
Transplantation-related mortality at 24 months[a] [%] 90% CI	[redacted]
Transplantation-related mortality at 36 months[a] [%] 90% CI	[redacted]
[a] Based on Kaplan-Meier estimates [Table 14.2.2A: Program: EBMT 2019 /SurvivalTRM /t_trm_fas]	

Summary results of overall survival (Full Analysis Set)	Treosulfan (N = 70)
Median follow-up[a] [months] (range of those surviving)	[redacted]
Subjects with event	[redacted]
Subjects without event	[redacted]
Overall survival at 12 months[b] [%] 90% CI	[redacted]
Overall survival at 24 months[b] [%] 90% CI	[redacted]
Overall survival at 36 months[b] [%] 90% CI	[redacted]
[a] Based on reverse Kaplan-Meier estimates [b] Based on Kaplan-Meier estimates [Table 14.2.3A: Program: EBMT 2019 /SurvivalOS_RFPFS_EFS /t_os_fas]	

Summary results of relapse/progression (Full Analysis Set)	Treosulfan (N = 70)
Subjects with event	[redacted]
Subjects without event Censored Death[a] Primary graft failure[a] Secondary graft failure[a]	[redacted]

Summary results of relapse/progression (Full Analysis Set)	Treosulfan (N = 70)
Cumulative incidence of relapse/progression at 12 months [%] 90% CI	[redacted]
Cumulative incidence of relapse/progression at 24 months [%] 90% CI	[redacted]
Cumulative incidence of relapse/progression at 36 months [%] 90% CI	[redacted]
[a] Only if this event occurred first [Table 14.2.4A: Program: EBMT 2019 /Survival RPS_NRM /t_cuminc_rps_summary_fas]	

Source: Company's response form to technical engagement

Appendix 5 Treatments total costs

Conditioning regimen	Dosing	Total cost (excluding wastage)
Treosulfan + Fludarabine (MAC)	Treosulfan: 10 g/m^2 once daily over 3 days prior to alloHSCT Fludarabine: 30 mg/m^2 once daily over 5 days prior to alloHSCT	£2,569.44
Busulfan + Fludarabine (RIC) - including Phenytoin	Busulfan: 4 x 0.8 mg/kg OR 1 x 3.2 mg/kg daily over 2 days prior to alloHSCT Fludarabine: 30 mg/m^2 once daily over 5 days prior to alloHSCT Phenytoin: 3 x 200 mg over 1 day then 3 x 100 mg over 3 days prior to alloHSCT	£1,796.33
Busulfan + Fludarabine (RIC) - including Phenytoin	Busulfan: 4 x 0.8 mg/kg OR 1 x 3.2 mg/kg daily over 2 days prior to alloHSCT Fludarabine: 30 mg/m^2 once daily over 5 days prior to alloHSCT	£1,794.70
Melphalan + Fludarabine (RIC) - including Alemtuzumab	Melphalan: Single dose of 140 mg/m^2 prior to alloHSCT Fludarabine: 30 mg/m^2 once daily over 5 days prior to alloHSCT Alemtuzumab: 20 mg/m^2 once daily over 5 days prior to alloHSCT	£2,601.90
Melphalan + Fludarabine (RIC) - excluding Alemtuzumab	Melphalan: Single dose of 140 mg/m^2 prior to alloHSCT Fludarabine: 30 mg/m^2 once daily over 5 days prior to alloHSCT	£901.89
Cyclophosphamide + TBI (MAC)	Cyclophosphamide: 60 mg/kg once daily over 5 days prior to alloHSCT TBI: 2 Gy fractions daily over 6 days (12 Gy total) prior to alloHSCT	£2,781.19
Cyclophosphamide + Busulfan (MAC)	Cyclophosphamide: 60 mg/kg once daily over 5 days prior to alloHSCT Busulfan: 4 x 0.8 mg/kg daily OR 1 x 3.2 mg/kg daily over 4 days prior to alloHSCT	£3,373.68
Cyclophosphamide + Thiotepa (MAC)	Cyclophosphamide: 60 mg/kg once daily over 5 days prior to alloHSCT Thiotepa: 5 mg/kg daily over 3 days prior to alloHSCT	£8,956.67
Busulfan + Fludarabine (MAC) - including Thiotepa	Busulfan: 3.2 mg/kg once daily over 3 days prior to alloHSCT Fludarabine: 50 mg/m^2 once daily over 3 days prior to alloHSCT Thiotepa: 5 mg/kg once daily over 2 days	£8,515.19
Busulfan + Fludarabine (MAC) - excluding Thiotepa	Busulfan: 3.2 mg/kg once daily over 3 days prior to alloHSCT Fludarabine: 50 mg/m^2 once daily over 3 days prior to alloHSCT	£2,612.47

Source: Company's response form to technical engagement