TECHNICAL REPORT FOR THE DRUG SAFETY AND EFFECTIVENESS NETWORK:
COMPARISON OF CONDITIONING REGIMENS FOR PATIENTS UNDERGOING ALLOGENEIC HEMATOPOIETIC STEM CELL TRANSPLANT: A SYSTEMATIC REVIEW AND META-ANALYSIS

Authors:

Dianna Wolfe, Fatemeh Yazdi, Brian Hutton, David Moher, Natasha Kekre, Chris Bredeson, David Allan

November 2016

Research Conducted by:

The Knowledge Synthesis Group:
Clinical Epidemiology Progam,
Ottawa Hospital Research Institute
501 Smyth Road
The Ottawa Hospital, General Campus
Center for Practice Changing Research
Ottawa, Ontario
Box 201b, K1H 8L6

EXECUTIVE SUMMARY

Background

Allogeneic hematopoietic stem cell transplantation (HSCT) has been used as a treatment for hematologic and lymphoid cancers since the 1960s, and continued improvements have been realized to refine HSCT as the best curative option for many patients. Healthy donor hematopoietic stem and progenitor cells can differentiate into all lineages of the lympho-hematopoietic system, giving rise to all the specialized cells of the blood and immune system. Prior to infusing donor cells, a patient's own hematopoietic cells are weakened or destroyed and the host immune system suppressed and/or eliminated to facilitate donor engraftment through the use of a series of chemotherapeutic drugs and/or irradiation procedures called a ***conditioning regimen***.

A variety of conditioning regimens have been developed over the years to achieve complete destruction of the hematopoietic and immune systems (myeloablative regimens) or their partial destruction (non-myeloablative or reduced-intensity regimens). Myeloablative regimens are associated with high non-relapse mortality in the first 100 days after transplantation due to organ toxicity and/or complications of immunosuppression (e.g., serious infection) prior to hematopoietic engraftment and immune reconstitution from the donor cells. However, myeloablative regimens are associated with a lower risk of relapse of the underlying cancer due to greater cytoreduction of the malignant cells. In non-myeloablative regimens and reduced intensity regimens, the lower doses have less impact on residual malignant cells but remain highly immunosuppressive, facilitating a more gradual takeover of the hematopoietic and immune systems by donor cells. While this approach reduces the risk of serious organ toxicity and early death due to transplant-related complications and allows older patients with co-morbidities to undergo HSCT, patients are still heavily immune-suppressed and remain at high risk of infections, rejection, and graft-versus-host disease.

The curative potential of non-myeloablative and reduced intensity conditioning therapies relies more heavily on the ability of the new graft-derived immune system to eliminate any residual primary cancerous cells not eradicated by the conditioning regimen—the "graft-vs-tumour effect"—which remains challenging to fully characterize and predict. Thus, while reducing the risk of death due to non-relapse causes in the pre-engraftment period, these regimens may be associated with an increased risk of relapse of the underlying disease. The ideal conditioning regimen would demonstrate a balance in the risks of both non-relapse mortality and relapse of the primary cancer. Comparing the effectiveness of conditioning regimens is ideally addressed by randomized controlled trials (RCTs) given the many factors that can influence outcomes following HSCT. A systematic search of RCTs addressing conditioning regimens in HSCT will allow us to identify optimal strategies to improve patient outcomes, identify standard treatment arms that should be used in future studies, provide insight on clinical outcomes that should be reported, and may permit network meta-analysis to infer potential comparisons between treatments that have not been directly compared before.

Objective addressed in this review:
To compare the benefits (e.g., reduction in mortality and relapse) and harms (e.g., increased risk of HSCT-related conditions) of competing regimens used to condition patients prior to undergoing HSCT, and to establish a hierarchy of intervention strategies according to their efficacy and safety.

Overview of Research Approach

Using data from RCTs, networks of evidence of conditioning regimens were developed. For regimens that have never been directly compared in head-to-head trials, their effects can be compared using ***network meta-analysis*** to derive comparisons between therapies and to rank all regimens according to their

relative effects on the outcomes of interest (e.g., the risks of mortality, relapse, and the development of several HSCT-related conditions such as graft-versus-host disease (GVHD), veno-occlusive disease (VOD), and bronchiolitis obliterans). This approach allows us to understand the strength of the evidence supporting the various conditioning regimens. This is helpful to generate recommendations for the optimal management of HSCT patients and for the selection of the most appropriate control arm in future RCTs to leverage the existing foundation of evidence.

Systematic Review Methods

The databases Medline, PubMed, Embase, and the Cochrane Register of Controlled Trials were searched for randomized controlled trials of patients undergoing HSCT. Studies were included if patients underwent allogeneic HSCT in the treatment of hematologic neoplasias or benign disease and were randomly allocated to receive a conditioning regimen. Outcomes of interest included overall mortality, non-relapse mortality, relapse of underlying disease, risk of acute and chronic GVHD, and specific harms. We conducted Bayesian network meta-analyses to compare conditioning regimens for outcomes of interest, where feasible. All outcomes were analysed as binary endpoints, with summary comparisons between regimens reported as odds ratios with 95% credible intervals. For outcomes for which network meta-analysis was not possible, detailed narrative summaries have been provided.

RESULTS

Eighteen trials assessed 18 unique conditioning regimens in 2,361 total patients; represented regimens are listed in **Table 1**. Three trials were not included in our network meta-analyses because either (1) they were conducted strictly on aplastic anaemia patients (2 studies, n = 213) or (2) the conditioning regimens evaluated were not described in detail (i.e., "standard conditioning" with or without total lymphoid irradiation; 1 study, n = 235). A total of 1,913 patients in 15 trials were available for inclusion in network meta-analyses. Overall, there was substantial variability in patient populations with respect to age, underlying hematologic disease, disease risk of relapse/mortality, and transplant donor status (i.e., related vs. unrelated, matched vs. unmatched). Trial publication dates ranged from 1988–2015. Networks of treatments were often disconnected, thereby limiting the ability to compare all interventions encountered.

Table 1: Interventions compared in trials of conditioning regimens for HSCT

Available for network meta-analyses:	***Not included in network meta-analysis:***
• CY+TBI • BU+CY • BU+FLUD • ATG+BU+FLUD • FLUD+TBI • ATG+TLI • TBI • ETP+TBI • MELPH+TBI	• CY (aplastic anaemia patients only) • CY+ATG+FLUD (aplastic anaemia patients only) • CY+ATG (aplastic anaemia patients only) • "Standard conditioning regimen" • "Standard conditioning regimen"+TLI
Disconnected from networks and summarized narratively:	
• ALZ+MELPH+FLUD • ATG+THIO+MELPH+CY • mBU+CY • mBU+FLUD	

ALZ = alemtuzumab; ATG = anti-thymocyte globulin; BU = busulfan; CY = cyclophosphamide; ETN = etanercept; FLUD = fludarabine; mBU+CY = BU+CY+ hydroxyurea+semustine+cytosine arabinoside; mBU+FLUD = BU+FLUD+hydroxyurea+ semustine+cytosine arabinoside; MELPH = melphalan; TBI = total body irradiation; TLI = total lymphoid irradiation

Results of network meta-analyses and narrative summaries

In the networks analysed, most comparisons between conditioning regimens were informed only by indirect evidence (i.e. head-to-head trials were not available), and many of the direct comparisons were based on single studies with small numbers of patients. Thus, a sparse evidence base and considerable between-study heterogeneity in patient populations complicated targeting of findings from meta-analyses. Separate network meta-analyses for overall mortality were conducted at the follow-up times of 100 days, and 1, 2, 4, and 5 years. Regarding disease relapse, only studies following patients for a median of 2–3 years were included in the network. CY+TBI was considered the standard conditioning regimen for comparison purposes. Clinical interpretation of findings were as follows:

- **Overall mortality:** 5 conditioning regimens were included in network meta-analyses at all follow-up times: CY+TBI, BU+CY, BU+FLUD, ETP+TBI, and MELPH+TBI. No statistically significant differences between regimens were identified at 100 days. However, CY+TBI demonstrated significantly reduced overall mortality at 1 year compared to BU+CY, and at 2 and 4 years, when compared to either BU+CY or BU+FLUD. Fewer patients contributed to overall survival outcomes estimated at 5 years of follow-up and no significant differences were found between regimens.

- **Non-relapse mortality (NRM):** Network meta-analyses could not be conducted due to reporting limitations on the competing risk of relapse. Narrative summaries demonstrated significantly reduced NRM at 1, 2, and 5 years for BU+FLUD compared to BU+CY in one study (Rambaldi, 2014), but not in 2 other studies that compared the same regimens (Lee, 2013 and Liu, 2013). At 1 year of follow-up, there was a trend toward FLUD+TBI offering reduced NRM compared to CY+TBI in one study (Bornhauser, 2011). Additionally, early and late (7 year) cumulative incidence of NRM was significantly reduced for CY+TBI when compared to BU+CY; however, when adjusted for early vs. advanced disease and donor age > or < 30 years in a multivariable model, no significant difference was found between the regimens (Ringden, 1999).

- **Relapse:** Four regimens provided data for a network meta-analysis of disease relapse at 2–3 years post-transplant (CY+TBI, BU+CY, BU+FLUD, and FLUD+TBI). There were no significant differences between regimens with respect to their effects on the risk of relapse.

- **Risk of acute and chronic GVHD:** All 9 regimens available (Table 1) were included in the network meta-analysis for aGVHD, while only 3 regimens (CY+TBI, BU+CY, and BU+FLUD) could be included for an analysis for cGVHD. For aGVHD, TBI alone was the top-ranked regimen; however, there was no significant difference in the risk of aGVHD when TBI was compared to any regimen other than ATG+BU+FLUD and MELPH+TBI. The standard conditioning regimen, CY+TBI, was was not significantly different from any other regimen. There were no significant differences in the 3 regimens included in the cGVHD network.

- **Additional harms-related findings:** A pairwise meta-analysis comparing CY+TBI to BU+CY for the risk of veno-occlusive disease (VOD) at 28 days post-transplant found no significant difference between the regimens. However, a network meta-analysis including CY+TBI, MELPH+TIB, BU+FLUD, and BU+CY, demonstrated that CY+TBI significantly reduced the risk of VOD compared to BU+CY at ≥100 days post-transplant. Other adverse events including specific organ toxicity and infections were described in a small number of studies, precluding meta-analysis. Based on available data, CY+TBI significantly reduced the risk of bronchiolitis obliterans when compared to BU+CY in long-term follow-up in one study. CY+TBI was associated with a higher risk of a positive blood culture at 100 days compared to BU+CY in one study; however, in another study, there was no significant difference between the two regimens in

the risk of bacteremia after 2 years. The single study comparing modified regimens of BU+CY and BU+FLUD identified a significantly higher risk of severe pneumonia in the modified BU+FLUD group after 1.4 years, which ultimately halted the study. Finally, the addition of FLUD to CY+ATG in conditioning regimens for aplastic anaemia patients was associated with significantly reduced regimen related toxicity and pulmonary complications.

- ***Key summary messages regarding conditioning regimens for HSCT were as follows***:

    - Inconsistency in reporting outcomes in RCTs of conditioning regimens precludes the ability to perform meta-analyses for several important transplant outcomes.

    - Network meta-analysis demonstrated that CY+TBI was associated with lower overall mortality at 1 year post-transplant compared to BU+CY, and at 2 and 4 years post-transplant compared to BU+CY and BU+FLUD.

    - Network meta-analysis did not identify statistically significant differences in any of the conditioning regimens with respect to relapse (4 regimens), acute GVHD (9 regimens), or chronic GVHD (3 regimens).

    - CY+TBI and BU+CY were associated with similar rates of VOD at 28 days post-transplant (2 studies); however, at ≥100 days post-transplant, CY+TBI significantly reduced the risk of VOD compared to BU+CY (4 studies)

    - CY+TBI was associated with reduced incidence of bronchiolitis obliterans compared to BU+CY (1 study).

Recommendations and Future Studies

Numerous drugs are used in a broad variety of single- and multi-agent conditioning regimens prior to HSCT. This review found that comparative evidence from randomized trials is lacking for many comparisons of these regimens, especially newer regimens, which has led to considerable practice variation between institutions. This systematic review of the evidence, incorporating network meta-analyses where possible, was conducted to address these gaps. Eighteen randomized trials of 18 conditioning regimens were identified and studied to inform analyses in this review.

The following key points for clinical practice were identified in this review:

- CY+TBI is associated with improved survival compared with BU+CY and with BU+FLUD

- The risks of relapse and acute and chronic GVHD were not significantly different between any regimens.

- CY+TBI is associated with a reduced risk of VOD after 100 days post-transplant compared with BU+CY

Future studies of conditioning regimens should be stratified for competing factors, such as patient age and underlying disease, and donor factors. A standard arm from the existing evidence network should be included in future studies to leverage existing knowledge. Consistent outcome reporting is urgently needed in HSCT studies to improve network analysis and comparison with other studies.

Supplemental Information: Definitions of relevance for readers of this summary

What is a systematic review?

- A systematic review is a type of literature review performed using a structured approach to identify, appraise and analyze research studies which are relevant for answering a question posed by a research team. Reviews of randomized controlled trials are commonly used for the practices of technology assessment and evidence based medicine. Findings from identified research studies can be synthesized in a variety of ways, one of which includes the practice of meta-analysis.

What is meta-analysis?

- Meta-analysis is a statistical technique that can be used to quantitatively combine the findings from research studies. For example, the odds ratios from a series of trials comparing rates of cGVHD incidence with two different conditioning regimens may be combined using meta-analysis to estimate a summary odds ratio based on all available evidence. Meta-analysis uses a weighted average approach to combine study data, and produces a precise estimate of the treatment effect of interest. Studies included in a meta-analysis are traditionally identified by a systematic review, and are used to compare two treatments of interest.

What is a network meta-analysis?

- A network meta-analysis can be thought of as a generalization of traditional meta-analysis where there is an interest in comparing more than two treatments. Studies included in a network meta-analysis are also traditionally identified by a systematic review process.

What is an odds ratio?

- An odds ratio is a measure of association between an intervention (or exposure) and an outcome. It represents the odds that an outcome will occur given a particular intervention, compared to the odds of the outcome occurring in the presence of a different intervention.

TABLE OF CONTENTS

ABBREVIATIONS

AA = aplastic anemia
aGVHD = acute graft-versus-host disease
ALL = acute lymphoblastic leukemia
AML = acute myeloid leukemia
Ara-C = cytosine arabinoside
ATG = antithymocyte globulin
BM = bone marrow
BU = busulfan
cGVHD = chronic graft-versus-host disease
CI = confidence interval
CML = chronic myelogenous leukemia
CMV = cytomegalovirus
CNS = central nervous system
CR = conditioning regimen
CrI = credible interval
CsA = cyclosporine A
CY = cyclophosphamide
DIC = deviance information criteria
DNA = deoxyribonucleic acid
ETP = etoposide
FE = fixed effects
FLUD = fludarabine
GVHD = graft-versus-host disease
HLA = human leukocyte antigen
HR = hazard ratio
HSCT = hematopoietic stem cell transplant
HU = hydroxyurea
ITT = intention to treat
K-M = Kaplan-Meier
MA = myeloablative
mBU+CY = BU+CY+ hydroxyurea+semustine+cytosine arabinoside
mBU+FLUD = BU+FLUD+hydroxyurea+semustine+cytosine arabinoside
MDS = myelodysplastic syndromes
Me-CCNU = semustine
MELPH = melphalan
MM = multiple myeloma
MMF = mycophenolate mofetil
MoAB = anti-P55 interleukin-2 receptor monoclonal antibody
MTX = methotrexate
NHL = non-Hodgkins Lymphoma
NMA = network meta-analysis
Non-MA = non-myeloablative
NR = not reported
NRM = non-relapse mortality
NSD = no significant difference
OR = odds ratio
PB = peripheral blood
PRISMA = Preferred Reporting Items for Systematic Reviews and Meta-Analyses

RCT = randomized controlled trial
RD = risk difference
RE = random effects
SUCRA = surface under the cumulative ranking curve
TAC = tacrolimus
TBI = total body irradiation
THIO = thiotepa
TLI = total lymphoid irradiation
UC = umbilical cord
UD = unrelated donor
VOD = veno-occlusive disease

1. BACKGROUND

1.1. Introduction

In the 1950s, animal models demonstrated that after low-dose total body irradiation, death could be averted by the transplantation of bone marrow from an identical twin (syngeneic) or from the animal's own (autologous) stored bone marrow[1]. With the discovery of human leukocyte antigens (HLAs) in the 1960s and their influence on graft survival, bone marrow transplants from genetically different individuals (allogeneic) that had matched HLA profiles became feasible, and allogeneic hematopoietic stem cell transfer (HSCT) became a viable potential treatment option for hematologic disease[1]. During more recent years, HSCT has been used to treat a variety of cancers and other diseases, including acute myeloid leukemia, acute lymphoblastic leukemia, chronic myeloid leukemia, juvenile chronic myeloid leukemia, chronic lymphocytic leukemia, myelodysplastic syndromes, myeloproliferative disorders, aplastic anaemia, and sickle cell anaemia, amongst other conditions[1]. Worldwide, the vast majority of HSCTs are used to treat lymphoid and hematologic cancers, with nearly half performed to treat acute leukemias.

In brief, allogeneic HSCT involves either complete or partial ablation of the hematopoietic system and profound immunoablation to facilitate engraftment of the donor blood and immune system, achieved through transplantation of healthy donor hematopoietic stem cells. These stem cells (1) produce daughter stem cells, so as to be self-renewing; and (2) differentiate to become the primitive progenitors of the entire lymphohematopoietic system that will give rise to specialized precursors of the various cell lineages.

Immediately prior to transplantation, patients have residual underlying disease and their cellular immunity eradicated through the administration of a *conditioning regimen*. Conditioning regimens can consist of a mixture of chemotherapy and/or total body irradiation and other agents, and can be prescribed in different strengths: *myeloblative*, *non-myeloablative,* and *reduced-intensity* regimens[2, 3]. While the intent of myeloablative regimens is to completely eradicate the patient's disease, non-myeloablative regimens eradicate only a portion of cancerous cells, and instead rely on the *graft-versus-tumor effect*—in which donor-derived alloreactive immune cells eliminate cancer cells over time—providing immune surveillance that can eliminate the remaining diseased cells. Because myeloablative regimens are associated with greater transplant-related mortality with increasing patient age, non-myeloablative regimens evolved to enable older and more medically fragile patients to undergo HSCT[2, 3]. However, non-myeloablative regimens are associated with a greater risk of relapse, due to the residual primary disease remaining following conditioning. Reduced-intensity regimens can be considered an intermediate category. The ideal conditioning regimen for a given patient will have a balance between the risks for relapse and mortality. Currently, the best choice of conditioning therapy remains unclear, and a robust systematic review is needed to address this knowledge gap.

1.1.1. Interventions available for use in conditioning regimens

A variety of interventions are used in conditioning regimens administered to patients undergoing HSCT. Generally speaking, when used together these agents work to destroy both healthy and diseased bone marrow, with the goals of eliminating disease and preventing graft rejection. Conditioning regimens primarily are formed of various pairings of the following components:

- ***Antithymocyte globulin (ATG)***: Antithymocyte globulin dramatically reduces the number of circulating T-lymphocytes through cell lysis. This reduces host immune response in the immediate post-transplant phase, but also increases the risk of infection.
- ***Busulfan (BU)***: Busulfan is an alkylating agent that selectively targets rapidly dividing cells of the bone marrow, including stem cells, acting through formation of intra-strand DNA crosslinks

that prevent DNA replication and cause cell death[4]. In high doses, it is myeloablative and can be used in place of total body irradiation. With other drugs, it is commonly used in conditioning regimens for patients with acute or chronic leukemias. Significant adverse effects can occur with its use due to highly variable pharmacokinetics. These effects have been partly reduced through greater use of intravenous rather than oral formulations. The most commonly used non-radiation-containing conditioning regimen is cyclophosphamide with busulfan.

- ***Cyclophosphamide (CY)***: Similar to busulfan, cyclophosphamide is an alkylating agent that causes intra-strand as well as inter-strand DNA crosslinks, preventing DNA replication and causing cell death. It preferentially targets lymphoid cells and is highly immunosuppressive. Associated important toxicities include hemorrhagic cystitis and secondary cancers.
- ***Etoposide***: Etoposide is a topoisomerase inhibitor that acts by rapidly dividing cancer cells to break DNA strands, inhibiting DNA synthesis, and promoting programmed cell death[5]. It can affect all cell lines in the bone marrow, leading to immunosuppression, anaemia, and thrombocytopenia, with their associated side-effects. It may be combined in high doses with total body irradiation as a conditioning regimen.
- ***Fludarabine***: Fludarabine is a purine analog that acts on ribonucleotide reductase and DNA polymerase to inhibit DNA synthesis, targeting both dividing and non-dividing cells[6]. It has relatively high specificity for both healthy and diseased blood cells, and causes suppression of all cell lines, resulting in immunosuppression, anaemia, and thrombocytopenia, with their associated side effects. Fludarabine is often used in reduced-intensity and non-myeloablative conditioning regimens.
- ***Melphalan***: Melphalan is an alkylating agent that causes inter-strand crosslinks in DNA, inhibiting DNA and RNA synthesis and causing programmed cell death[7]. In conditioning regimens, it can be used at myeloablative or reduced-intensity dosages, with or without other agents.
- ***Thiotepa***: Thiotepa is an alkylating agent that causes inter-strand crosslinks in DNA, inhibiting DNA and RNA synthesis and causing programmed cell death[8]. It was recently designated for use in conditioning regimens and in high doses is myeloablative.
- ***BEAM regimen (carmustine (BCNU), etoposide, cytarabine, melphalan)*:** The BEAM regimen is a combination therapy that has fewer side-effects than widely used cyclophosphamide/total body irradiation regimens. The BEAM regimen alone may not be sufficiently immunosuppressive to allow allogeneic stem cell engraftment and is used preferentially in autologous HSCT to treat lymphoma. If combined with low-dose total body irradiation, fludarabine or antithymocyte globulin therapy, it can be considered for allogeneic transplantation for patients with lymphoma.
Total body irradiation (TBI): As the name suggests, total body irradiation is the exposure of the entire body to ionizing radiation, with the goal to destroy bone marrow and cancer cells[9]. Typically, the full dose of radiation is divided into smaller doses (fractionated) given over several days, to reduce toxicity and increase tolerability. It is commonly combined with chemotherapy agents, including cyclophosphamide, etoposide, or cytarabine. Major side-effects include mucositis, lung toxicity, and female infertility.
- ***Total lymphoid irradiation (TLI)***: Total lymphoid irradiation is the selective exposure of all major lymph nodes, thymus, and spleen to ionizing radiation, while the non-lymphoid organs are shielded. It causes a potent and long-lasting immunosuppression due to reduced total lymphocytes and T cells, with fewer of the severe side-effects associated with total body irradiation.

Conditioning regimens can be broadly categorized into three groups by their ability to remove bone marrow: *myeloablative*, *non-myeloablative*, and *reduced intensity* regimens[2]. There is interest in comparing the effectiveness and safety of these groups to determine which may provide the greatest balance of benefits and harms for patients. The groups are described as follows:

- ***Myeloablative***: Myeloablative regimens destroy bone marrow, lymphoid tissue, and cancer cells such that hematologic recovery cannot occur. The combinations of BU/CY and CY/TBI are both classified as myeloablative conditioning regimens. Other pharmaceuticals belonging to the myeloablative grouping include etoposide, thiotepa and melphalan. Myeloablative regimens can be associated with high transplant-related mortality, which is influenced by factors such as patient and donor ages, severity of disease, patient co-morbidities and HLA matching.
    - Working definition: A combination of agents expected to produce profound pancytopenia (i.e., profound reduction in the numbers of all circulating blood cell types and platelets) and myeloablation (i.e. elimination of bone marrow activity, also associated with profound reduction in the numbers of all circulating blood cell types and platelets) within 1–3 weeks of administration; pancytopenia is long lasting, usually irreversible and in most instances fatal, unless hematopoiesis is restored by HSCT.

- ***Non-myeloablative***: Non-myeloablative regimens were introduced to reduce toxicity relative to myeloablative regimens, allowing patients of greater age and/or increased medical fragility to undergo HSCT. These regimens cause minimal cytopenia (i.e., reduction of circulating blood cells), but have sufficient immunosuppressive effect that full engraftment usually occurs following transplant of allogeneic blood stem cells. This class of regimens is commonly described as being "immunoablative" due to their strong immunosuppressive effect. Transplant-related mortality is typically reduced with this category of regimens in comparison with myeloablative regimens.
    - Working definition: A regimen that will cause minimal cytopenia and does not require stem cell support.

- ***Reduced intensity regimens***: Reduced-intensity regimens are considered an intermediate category of conditioning regimens that do not fit the definition of myeloablative or non-myeloablative. They differ from myeloablative regimens in that doses of agents are reduced by a minimum of 30% and transplant-related mortality is reduced. They differ from non-myeloablative regimens in that they cause significant pancytopenia and require stem cell support. That said, hematopoietic recovery without allogeneic transplantation may be possible eventually, but the time required would result in significant morbidity and mortality due to prolonged myelosuppression and profound immunosuppression. Fludarbine is commonly used in combination with either an alkylating agent at a reduced dose (e.g., busulfan, melphalan or thiotepa) or reduced dose TBI.
    - Working definition: a regimen that cannot be classified as myeloablative or non-myeloablative.

1.1.2. Why this review is important

Numerous agents are used in a broad variety of conditioning regimens for patients undergoing allogeneic HSCT. Additionally, considerable variation in practice exists between institutions, given the lack of head-to-head comparative evidence to support these interventions. A systematic review of the evidence, incorporating network meta-analyses to compare regimens and their impact on key outcomes, will help to explore the relative benefits and harms of competing interventions. Identifying regimens that best balance benefits and harms will improve patient outcomes.

1.1.3. Objective of this systematic review

To compare the benefits (e.g., reduction in of mortality and GVHD) and harms (e.g., increased risk of relapse) of competing regimens used to condition patients prior to undergoing HSCT, and to establish a hierarchy of intervention strategies according to their efficacy and safety.

2. REVIEW METHODS

The checklist from the Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) Extension Statement for Network Meta-analyses[10] was used to guide preparation of this report. A completed checklist is provided in the report's appendices.

2.1. Protocol and Registration

The protocol for the review was developed using an integrative framework amongst researchers from the Ottawa Hospital Research Institute, the Canadian Bone Marrow |Transplant Group and the pan-Canadian Oncology Drug Review, knowledge. The protocol for the review was subsequently made openly available online in the University of Ottawa Library's online repository, and can be found at the following link: https://www.ruor.uottawa.ca (search the term 'DSEN stem cell transplantation'). The protocol was also registered in PROSPERO (CRD42015026738).

2.2. Study Eligibility Criteria

Table 1 provides a detailed summary of the eligibility criteria set out for this review using the Population-Interventions-Comparators-Outcomes-Study design (PICOS) framework.

Table 1: Overview of study eligibility criteria

PICOS Element	Description
Population	Patients undergoing allogeneic HSCT for the treatment of Acute Lymphoblastic Leukemia (ALL), Non-Hodgkins Lymphoma (NHL), Chronic Myelogenous Leukemia (CML), Myelodysplastic Syndromes (MDS), Acute Myeloid Leukemia (AML), Multiple Myeloma (MM), Aplastic Anaemia (AA), or other cancers.
Interventions/Comparators	Regimens involving combinations of cyclophosphamide, busulfan, melphalan, fludarabine, total body irradiation, etoposide, BEAM, ATG, alemtuzamab and lymphoid radiation will be included. Myeloablative, non-myeloablative, and reduced intensity regimens were all eligible. Additional regimens were incorporated as needed following identification of eligible studies. As there is no definitive source of relevant doses of these interventions to consider for comparisons, studies involving any dose of agents will be retained during study identification, and we will review intervention details with clinical experts to determine necessary dose stratifications and/or exclusions of specific studies involving irrelevant regimens.
Outcomes	Overall mortality Transplant-related mortality Relapse of underlying disease Acute and chronic GVHD Veno-occlusive disease (VOD) (known also as sinusoidal obstruction syndrome) Bronchiolitis obliterans.
Study Design	Randomized controlled trials with at least 100 days of patient follow-up following transplant.

- ***Dealing with duplicate publications and other characteristics.*** For studies that are associated with multiple publications (e.g., updates of different follow-up durations), we will retain the most up-to-date reports and make note of all related manuscripts. Only studies published in English will be retained for inclusion[11].
- ***Potential for grouping of eligible interventions in the treatment network.*** The following comparisons of and/or groupings of treatments were considered in potential sensitivity analyses with regard to constructing the geometry of the treatment networks and how interventions may be grouped if deemed helpful by clinical experts: low-dose radiation regimens (<500 cGy) compared with high dose radiation regimens (>500 cGy), antibody-containing regimens (e.g., ATG)

compared with similar regimens without antibody therapy, and fludarabine-containing non-myeloablative regimens. We also explored the feasibility of a sensitivity analysis wherein conditioning regimens from the included studies were categorized according to intensity of the regimen as outlined in **Table 2**; these analyses were found to be infeasible because the conditioning regimens within most studies were consistently from the same intensity grouping. Preliminary classification of anticipated regimens to be identified is outlined in **Table 2**; additional regimens were classified and incorporated.

Table 2: Classification scheme for conditioning regimens

Myeloablative Regimens	Non-Myeloablative Regimens	Reduced Intensity Regimens
CY+TBI	CY+FLU±ATG	FLU+MELPH
BU+CY+TBI	FLU+AraC+Ida	BU+FLU
BU+CY	FLU+TBI (≤500cGy)	FLU+AraC
BU+CY+ATG	TBI (≤500cGy)	
BU+MELPH	Cladribine + AraC	
	TLI+ATG	

AraC = cytosine arabinoside; ATG = anti-thymocyte globulin; BU = busulfan; CY = cyclophosphamide; FLU = fludarbine; MELPH = melphalan; TBI = total body irradiation; TLI = total lymphoid irradiation

2.3. Search strategies to identify relevant literature

In June 2013, requestors of this DSEN query conducted preliminary work with members of the MAGIC team based at The Ottawa Hospital to explore some of the literature available for developing clinical guidance related to allogeneic HSCT. This was conducted in the form of a scoping review of published RCTs, and involved a systematic search for studies for various aspects of care in the realm of HSCT. The search was developed and conducted with the input of an information specialist and covered the following databases: Medline, PubMed, Embase, and the Cochrane Register of Controlled Trials. The search was also peer reviewed by a second information specialist using PRESS criteria[12]. The search strategy is provided in **Appendix 1**, along with a flow diagram summarizing results from screening.

Following screening of abstracts and then potentially relevant full text reports by two independent researchers, approximately 700 RCTs were identified in relation to the following aspects of care: 1) donor selection and source of cells, 2) conditioning regimens, 3) prevention and treatment of GVHD, 4) transfusion-related interventions, 5) prevention and treatment of infections, 6) prevention and treatment of hepatic sinusoidal obstruction syndrome, 7) prevention and/or treatment of bronchiolitis obliterans, and 8) others that were not otherwise classifiable. There were approximately 50 RCTs that were identified as relevant to the comparison of conditioning regimens. We updated the search to identify new studies published since June 2013 for inclusion in the proposed review in August 2015, producing a total of approximately 2,000 additional citations for review. These were combined with those from the initial search to establish our evidence base.

Appendix 1 presents the complete set of search strategies for the different databases searched.

2.4. Process of study selection

For new citations obtained from the updated search, review of citations based on title, keywords, and abstract (Level 1 screening) and full text articles (Level 2 screening) were carried out independently by two reviewers. Level 1 citations deemed potentially relevant or lacking sufficient information to make a decision were carried forward to Level 2. Study selection was conducted using Distiller Systematic Review Software (DSR) (Evidence Partners, Inc; Ottawa, Canada). Where consensus was not achieved following discussion, a third independent party was consulted to settle disagreements. At both stages of screening, a pilot exercise of a number of abstracts/full texts was performed to establish a baseline amongst the reviewers. The process of literature selection has been reported using a flow diagram as

recommended by the PRISMA statement[13], and encompasses both the 2013 and 2015 searches performed. Studies were not screened on outcome; however, studies included at the full-text stage that did not have an outcome of interest to the review did not move forward for data extraction.

2.5. Data collection from included studies

Primary data collection of included studies was performed independently by two reviewers using a standardized electronic data collection form in DSR. Collected data was compared for accuracy and agreement, with disagreements settled by discussion. The following elements were collected for each included study:

- study characteristics (authors, year of publication, journal, countries of performance);
- patient characteristics (eligibility criteria, number per group, and key demographics including age, gender, primary disease diagnosis, disease duration, comorbidities, HLA and CMV matching, etc.);
- Conditioning regimen data (drug(s) and radiation used as well as dosage, reported intensity category (reduced intensity, myeloablative or non-myeloablative) and other aspects of administration noted within each study);
- Outcome data (number of events and number of patients randomized for binary endpoints, and means with standard deviations for continuous endpoints).

To standardize follow-up times for mortality analyses, overall mortality data were extracted from Kaplan-Meier (K-M) survival curves at 100 days, as well as at 1, 2, 4, and 5 years of follow-up using data digitization software (Digitizelt, © I. Bormann 2001–2016; http://www.digitizeit.de).

All study characteristics were summarized in tabular form to facilitate inspection and discussion with clinical experts in terms of study heterogeneity, grouping of interventions, and other such topics required to inform analysis. Two review team members (DA, NK) also categorized the conditioning regimens in the included studies by intensity.

All relevant RCTs were evaluated using the revised Cochrane risk-of bias (RoB) tool[14]. The Cochrane RoB tool evaluates seven domains including sequence generation, allocation concealment, blinding, missing outcome data, selective outcome reporting, attrition, and "other sources of bias." Other sources of bias evaluated for this review included dissimilarities between treatment groups, with respect to demographics, comorbidities, co-interventions, and other factors. An overall assessment of the RoB for each study was determined for each general outcome category (i.e., mortality, GVHD prevention, and harms). Any disagreements were resolved through discussion. Results from these appraisals have been summarized in the appendices to the report.

2.5.1. Outcomes of interest

2.5.1.1. Mortality

Mortality outcomes of interest included overall mortality at 100 days as well as at 1, 2, 4, and 5 years post-transplant. Because these data were digitally extracted from K-M curves at the time points of interest, follow-up time was considered consistent between studies.

2.5.1.2. Acute GVHD

Based on input from participating clinical experts, the aGVHD outcome was limited to Grades II–IV aGVHD, based on current scoring systems. Studies that reported only overall aGVHD, with no limits on grade, were excluded from the review.

2.5.1.3. Chronic GVHD

The development of cGVHD generally occurs after day 100 post-transplant. The risk of mortality in the first 100 days post-transplant is generally high, meaning that many patients initially randomized in a study may not survive to the commencement of the cGVHD follow-up period. To account for the competing risk of death in the first 100 days, the most unbiased denominator for cGVHD analyses was considered to be only the patients alive at 100 days post-transplant (as opposed to the number originally randomized). When the number of patients alive at 100 days post-transplant was not reported in the text, where possible, it was derived from K-M survival curves.

Amongst included studies, cGVHD was reported either as overall cGVHD or scored as extensive or limited and reported by these grades. Based on input from the research team's participating clinical experts, the risk of extensive cGVHD was chosen as the preferred outcome for the evaluation of cGVHD prophylaxis. Thus, the preferred data for the cGVHD outcome was considered to be the risk of extensive cGVHD in patients that survived to day 100 post-transplant.

Based on expert opinion, the risk of cGVHD was considered highest in the first 2 years post-transplant. Thus, where data were available for multiple follow-up times in a study, the follow-up period most inclusive of the 2-year time point was preferred over early follow-up or long-term follow-up data.

2.5.1.4. Relapse

Based on expert opinion and in considering the available study data, the ideal follow-up time for analysis of the relapse outcome was determined to be between 2–3 years of follow-up. Only studies reporting data for this time period were used for network meta-analysis. Other follow-up times have been summarized narratively.

2.5.1.5. Specific harms

Upon consultation with the team's clinical experts, the following specific harms of interest were identified: total adverse events, severe adverse events, sepsis, pneumonia, CMV infection, fungal infection, other infections, renal toxicity, hepatic toxicity, veno-occlusive disease (VOD), and other harms (e.g., multi-organ failure, neurotoxicity). Only VOD had sufficient data available for network meta-analysis and the other harms have been summarized narratively.

2.6. Classification of interventions for network meta-analysis

Input from clinical expert opinion deemed that analyses for this review should be focused at the regimen/drug level, with no additional reflection of specific regimen intensities when establishing treatment groups for meta-analyses. Thus, all conditioning regimens using the same drugs were grouped into the same node in the network, regardless of intensity. Ultimately, there was no variation in intensity of regimens within the treatment nodes.

2.7. Methods for meta-analysis and network meta-analysis

A priori, there was interest to explore the feasibility of network meta-analysis for all clinical endpoints given the presence of multiple interventions of interest. For each outcome, we first assessed whether a connected treatment network of interventions was present (a necessary criteria for NMA). When this was present, we next explored the extent of homogeneity/similarity of patient populations and study methods across included studies (including consultation with the team's clinical experts), as the performance of valid network meta-analyses requires a similar distribution of effect modifiers across studies; where more than one study was present for pairwise comparisons within the treatment network, we also assessed statistical heterogeneity of corresponding treatment effects using the I^2 statistic. If homogeneity within pairwise comparisons and across comparisons within the network were judged sufficient, network meta-

analyses were performed. Where studies could not be included in a network due to poor reporting of specific interventions or extensive between-study heterogeneity was present, a narrative summary of the study was undertaken. Similarly, where networks could not be generated due to fragmentation and low numbers of studies, pairwise ORs were calculated to identify regimen comparisons with significant associations with the various outcomes, and narrative summaries for each outcome were written.

Traditional pairwise frequentist meta-analyses were also conducted using the statistical software Comprehensive Meta-analysis© (Biostat, Inc., Englewood, NJ: https://www.meta-analysis.com/). Both fixed and random effects models were fit and assessments of heterogeneity were made using the I^2 statistic. The results of the traditional pairwise meta-analyses were useful in the evaluation of heterogeneity and inconsistency in the network meta-analyses discussed below.

Bayesian network meta-analyses were conducted using well established methods described by the National Institute for Clinical Excellence (NICE)[15-17]. All analyses were conducted using WinBUGS software and the algorithms presented in technical guidance documents published by NICE. Both fixed and random effects models were fit to arm-level data. All models were fit using 10,000 or more burn-in iterations and 10,000 sampling iterations. Model convergence was assessed by inspection of trace plots and Monte Carlo standard error of the estimated parameters. Model fit was evaluated by posterior residual deviance values for both FE and RE models, with the model fit assumed to be adequate if the posterior residual deviance was similar to the number of data points in the model. DIC values were used to compare the relative fit of the models—a model was identified as having a relatively better fit if its DIC value was 5 points or lower than the DIC of another model. However, if the network contained many single-study connections between interventions, results from the FE model were preferred. All of the networks analysed for this report were comprised of many single-study connections; thus, results from FE models are focused on for summary of findings while findings from RE models are provided in the report's supplement. This approach has previously been seen in applications in other clinical realms and avoids misleading summary estimates which can be a consequence of the presence of a poorly estimated between-study variance parameter.[18]

2.8. Summarizing measures of effect

2.8.1. Summary Measures and Reference Intervention

All outcomes of interest were assessed as binary endpoints, allowing pairwise comparisons (whether from individual studies or from meta-analysis) to be summarized using odds ratios (ORs) and corresponding 95% CrIs. Odds ratios obtained through traditional pairwise meta-analyses were compared to ORs obtained from network meta-analyses to aid in evaluation of consistency between the direct and indirect evidence of the network.

For each outcome of interest, forest plots were generated to provide a visual display of efficacy of each regimen in the network compared to the current treatment standard. Input from our clinical experts identified the current conditioning regimen standard as cyclophosphamide plus total body irradiation CY+TBI). All ORs <1 in analyses presented in this report suggest the comparator regimen had greater efficacy than CY+TBI, while ORs >1 indicated the comparison regimen had lower efficacy than CY+TBI. For all analyses, 95% credible intervals which included the value of 1 were considered to indicate that there was no statistically significant difference between the comparison regimen and CY+TBI.

2.8.2. Graphical Presentations of Findings

Forest plots of summary comparisons versus the reference intervention (CY+TBI) have been presented for all endpoints where network meta-analyses were performed. Additionally, *league tables* presenting all

pairwise comparisons estimated from network meta-analysis were prepared; in these tables, regimens in the upper/left-most region of the league table have potentially greater efficacy than regimens appearing lower and further right in the table. For each league table, interventions are ordered from left to right in terms of decreasing SUCRA value, meaning preferred interventions are those presented on the left of the table (as SUCRA values nearer 1 suggest preferred interventions). The SUCRA value[19] for each regimen has also been presented above the regimen, allowing evaluation of relative rankings of regimens. One regimen may appear above another regimen in the league table; however, their SUCRA values may show little difference, indicating inferences regarding relative ranking should be made with caution. ORs and 95% CrIs should be considered the primary means of assessing the importance of differences between regimens.

League tables can be complex to interpret in the presence of many interventions, a situation encountered in the analyses presented in this review. Furthermore, when evidence networks consist of many interventions and the comparisons made within trials are both broad, there may be varying degrees of faith in pairwise comparisons dependent upon the number of intermediate treatments between therapies; those with just one intermediate therapy are often called *simple indirect comparisons*, while those with more intermediate therapies are called *compound indirect evidence.* Comparisons of treatments involving direct evidence (i.e., where at least one head-head trial informs the comparison of two therapies) are commonly considered of greatest validity, while comparisons informed by simple indirect evidence are typically considered of greater validity than comparisons informed by compound indirect evidence. Due to the diverse nature of the evidence networks in this review which consist of several comparisons based on compound indirect evidence, we have color-coded league tables of summary findings to demonstrate key sources of evidence for each comparison as follows:

- Red squares denote pairwise comparisons with at least one head-to-head trial available;
- Orange squares denote pairwise comparisons with simple indirect evidence available;
- Yellow squares denote pairwise comparisons with compound indirect evidence available.

2.8.3. Assessment of heterogeneity and inconsistency for network meta-analysis

An important step in the practice of systematic reviews that incorporate network meta-analyses is the validation of the assumption that patients in the included trials are 'jointly randomizable,' or in other words, that they are sufficiently homogeneous clinically that a patient in any one of the studies could have been a patient in any of the other included trials.[20] We empirically evaluated this assumption by review of the patient eligibility criteria and pertinent patient demographics, in collaboration with our participating clinical experts (DA, NK).

To assess the presence of clinical and methodologic heterogeneity within each pairwise comparison of the treatment network, we inspected trial and population characteristics. This was performed by inspection of tabulated lists of descriptive statistics for the a priori characteristics mentioned below (i.e., means and frequency distributions as appropriate for each characteristic), as well as review of measures of statistical heterogeneity mentioned above. Specifically, differences in the distribution of patients across primary disease categories (e.g., ALL, AML, CML) for the 3 conditioning regimens with the largest numbers of patients (CY+TBI, BU+CY, and BU+FLUD) were evaluated with a χ^2 test. Partitioning of the standardized χ^2 residuals was conducted to identify any disease/intervention combinations with observed numbers of patients that were significantly different than expected. Content experts were consulted to determine if any of these statistically significant differences were clinically relevant.

To ensure homogeneity and similarity across pairwise comparisons in the treatment network, we compared the descriptive statistics of key measures across the different pairwise comparisons in the network to verify they were similar. To identify covariates necessary for review, we consulted our clinical

expert team members and grouped traits that were identified in past studies of prognostic risk factors[21-24]. The following characteristics were considered most important to the establishment of transitivity within and across pairwise comparisons:

- Average patient age, gender distribution, and race distribution;
- % of patients with different primary diseases (e.g., AML, ALL, CML);
- Disease risk (based on clinical experts' assessments);
- Presence of comorbidities (pulmonary disease, rheumatologic disease, renal dysfunction, etc.), if reported;
- Year of study publication (for consideration of changes in co-interventions such as newer antimicrobials, molecular screening tests for infection, and GVHD prophylaxis regimens used);
- % of patients receiving an unrelated donor transplant;
- % of patients with full HLA match;
- % of patients with mismatched donor gender;
- % of patients that were cytomegalovirus (CMV)-seronegative and received transplants from CMV-seropositive donors;
- source of donor cells (bone marrow, peripheral blood stem cells, umbilical cord blood).

Another key assumption underlying NMA is that of *consistency*. That is, there is no conflict between direct and indirect evidence[17] that could result from heterogeneity in effect modifiers in the studies contributing to the direct and indirect evidence. To assess consistency, one commonly compares DIC statistics in fitted consistency and inconsistency models. As well, the posterior mean deviance of the individual data points in inconsistency models may be plotted against the posterior mean deviance in consistency models to identify any loops in the network where inconsistency was present. Additionally, NMA estimates may be qualitatively compared with direct frequentist pairwise estimates. There were no closed loops in any of the networks analysed for this review; thus, consistency between direct and indirect evidence was not an issue.

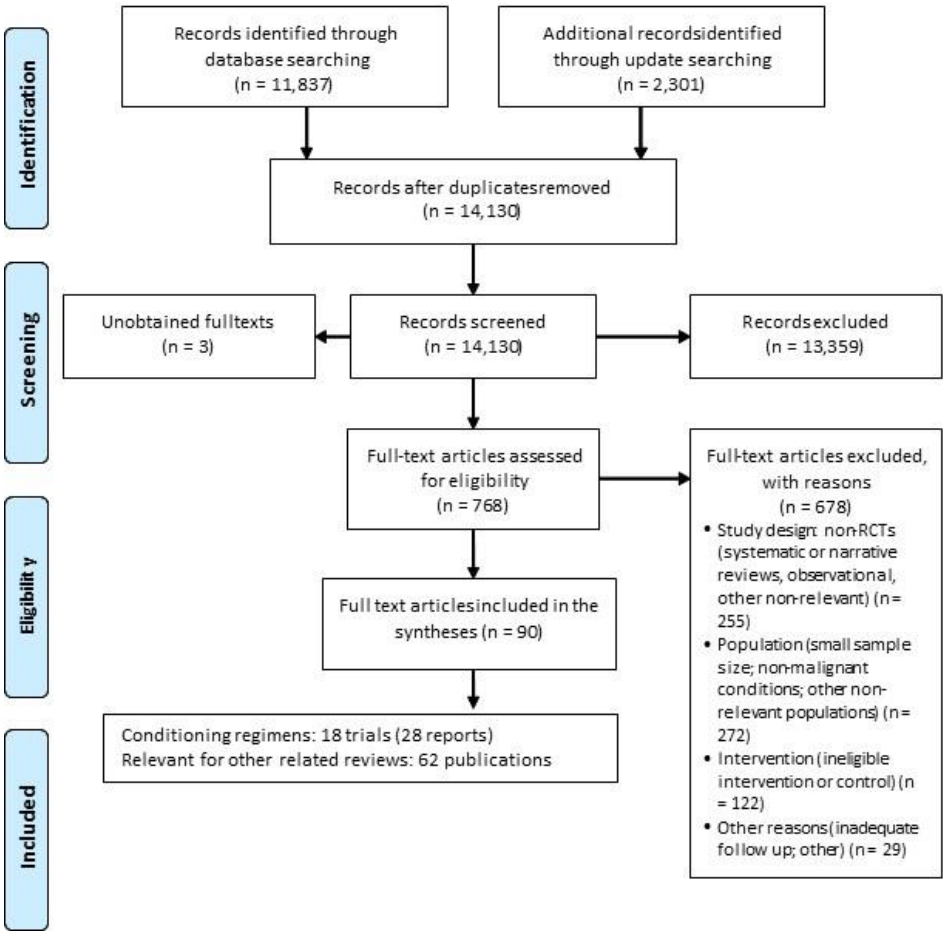
Although planned in the review's protocol to further establish the robustness of findings from primary analyses, subgroup analyses and meta-regression analyses were not feasible due to the presence of many single-study connections in most evidence networks, as well as a failure of studies to report outcomes in patient subgroups of relevance. This is further discussed later in the report.

3. REVIEW FINDINGS

3.1. Availability of relevant literature

Figure 1 presents a summary of the process of study selection. The electronic literature search identified a total of 14,130 citations for review following removal of duplicates. Stage 1 screening identified a total of 768 citations that were considered to be potentially eligible, and which were subsequently retrieved in full text for Stage 2 screening. Following screening of full text articles, a total of 28 publications describing 18 unique studies were retained (studies for 2 ongoing related reviews were also retained).

Figure 1: PRISMA flow diagram of study selection



3.2. Overview of study characteristics

Originally 19 studies were included. Data for one study [25] could not be used because the sample in each randomized group was not clearly stated and therefore the total number of trials presented in this review is 18, reported in 28 publications[26-51]. A total of 2,355 patients were included in all trials, while 1,913 patients contributed data to network meta-analyses. Median year (range) of publication was 2012 (1988–2015). Totals of 1 (6%), 6 (33%), 1(6%), and 10 (56%) trials were initially published between 1980–1990, 1991–2000, 2001–2010, and after 2010, respectively. Twelve studies (67%) were conducted in multiple sites, two (11%) in single sites, and this information was not reported for the remaining 4 studies (22%). Funding for the included trials was received from both industry and government/not for profit in 2 (11%), and 11 (61%), respectively. No information regarding the financial support was reported in the remaining 5 studies (28%). In all studies, patients were followed for at least one year (range 1–9 years).

3.3. Overview of patient characteristics

Eight studies (44%) enrolled patients $\leq$12 years of age as part of the inclusion criteria (range 1–55). Age range in the remaining 10 studies was 17–65 years. The median proportion of female patients was 37.3% (range 24–52%) in the set of 15 studies (83%) that reported this data. Information regarding donor source was reported in 17/18 studies; donor source was 100% bone marrow in 7 studies (39%), and peripheral blood in 3 studies (17%). In two studies (11%) a minority (1–2%) of transplants were from umbilical cord

blood[34,45]. Full HLA matching for all patients was used in 12 studies (67%), while it was possible for 60–92% of participants in four studies[33, 43, 45, 47], and not reported in two studies[26, 48]. The range of proportions of patients that received female-to-male transplants was 14–36% in 8 studies (44%). The remaining 10 studies (56%) did not report sufficient information regarding gender matching (**Table 4**).

All studies described the distribution of patients' underlying hematologic diseases, however limitations in reporting were identified. One study[51] reported the % of patients with lymphoma, without reporting the distribution of Hodgkin and non-Hodgkin lymphoma. In this case, patients with both diseases were assumed to have been included in the 4 patients classified as having lymphoma. Two studies[35, 37] were conducted in 100% CML patients, with both studies comparing CY+TBI to BU+CY. Six studies were conducted only in AML patients[26,29,33,41,47,48]. Three of these studies compared CY+TBI to BU+CY, FLUD+TBI, or MELPH+TBI, while the fourth study compared BU+CY to BU+FLUD. Studies that included only aplastic anaemia patients[34, 43] were not included in the analyses but have been summarized narratively. An appendix reports the distribution of underlying disease in all included studies, while **Table 3** provides the proportion of patients having each underlying disease for the three main regimens in the network meta-analyses to follow. Of note, there were statistically significant differences in the frequency distribution of the underlying diseases across the three regimens (χ^2 = 101.1827, degrees of freedom = 8, $p < 0.001$). Cells in bold text in **Table 3** identify where the observed number of patients evaluated was significantly different than expected, assuming a χ^2 frequency distribution.

Table 3: Proportion of patients with underlying diseases for the three main regimens in the NMAs

Underlying disease	CY+TBI (n = 380)	BU+CY (n = 457)	BU+FLUD (n = 115)
Aplastic anaemia	0%	0%	0%
Acute lymphoblastic leukemia	**4.7% (n = 18)[a]**	**14.9% (n = 68)[b]**	**20% (n = 23)[b]**
Acute myelogenous leukemia	55.0% (n = 209)	**43.1% (n = 197)[a]**	**76.5% (n = 88)[b]**
Chronic myelogenous leukemia	**39.7% (n = 151)[b]**	**40.7% (n = 186)[b]**	**0.9% (n = 1)[a]**
Myelodysplastic syndrome	**0% (n = 0)[a]**	**0.9% (n = 4)**	**2.6% (n = 3)[b]**
Multiple myeloma	0%	0%	0%
Non-Hodgkin's lymphoma	0.5% (n = 2)	0.4% (n = 2)	0% (n = 0)
Non-malignant disease	0%	0%	0%
Other malignant disease	0%	0%	0%

[a]The number of patients evaluated with this underlying disease is statistically significantly less than expected for this intervention, using a chi-squared test for frequency; [b]The number of patients evaluated with this underlying disease is statistically significantly more than expected for this intervention, using a chi-squared test for frequency; BU = busulfan; CY = cyclophosphamide; FLUD = fludarabine; TBI = total body irradiation

Assessment of disease risk based on the methods described earlier found that (1) 4 studies[26, 29, 41] provided insufficient information to establish the proportion of high risk patients; (2) no studies enrolled 100% high risk patients, while 5 studies[34, 35, 37, 40, 43, 48] enrolled none; and (3) amongst the remaining studies, the enrolled proportion of high risk patients ranged from 7 to 56% (**Table 4**). Findings from these evaluations highlight both limited information reported regarding study populations in some cases as well as a diverse range in populations across studies.

Table 4: Summary of characteristics of studies included in the systematic review of conditioning regimens for allogeneic HSCT											
Author (year)	**# of patients**	**Conditioning regimens compared**	**Intensity of Conditioning Regimens**	**Patient age range (years)**	**Donor source (%BM/PB/ UC/BM+PB)**	**% related donor**	**% full HLA match**	**% female donor / male recipient**	**% high risk**	**GVHD prophylaxis**	**Endpoints reported**
Baron 2014/ 2015[27, 28]	94	FLUD+TBI; ATG+TLI	Non-MA; Non-MA;	32–73	0/100/0/0	57	100	22	27	MMF+TAC	∞^{e}, $\S^{d}$, α^{c}, β^{e}, ε^{d}, ϕ^{d}, ρ^{d}, θ^{d}
Rambaldi 2014/ 2015[38, 48]	245	BU+FLUD; BU+CY	MA; MA	40–65	31/69/0/0	46	NR	NR	0	CsA+MTX	∞^{c}, $\S^{d}$, ε^{c}
Algarotti 2013[26, 52]	112	ALZ+MELPH+FLUD; ATG+THIO+MELPH+CY	Non-MA; Non-MA	17–67	NR	0	NR	NR	NR	NR	$\S^{d}$, α^{f}, β^{f}, ε^{cf}
Blaise 2013[31]	139	ATG+FLUD+BU; FLUD+TBI	Non-MA; Non-MA	21–65	0/100/0/0	100	100	36	24	CsA to both groups; MMF to FLUD+TBI group	∞^{e}, $\S^{d}$, α^{c}, β^{ce}, ε^{d}
Kornblit 2013[44]	85	FLUD+TBI; TBI	Non-MA; Non-MA	17–73	0/100/0/0	100	100	35	38	CsA+MMF	∞^{e}, $\S^{d}$, α^{c}, β^{f}, ε^{cf}, ϕ^{d}, θ^{d}, π^{d}
Lee 2013[45]	126	BU+CY; BU+FLUD	MA; MA	17–59	44/54/2/0	76	92	14	7	CsA or CsA+MTX	∞^{c}, $\S^{d}$, α^{c}, β^{c}, ε^{c}, μ^{c}, λ^{d}, ρ^{d}
Liu 2013[46]	105	BU+FLUD+ (HU+Me-CCNU+Ara-C); BU+CY+ (HU+Me-CCNU+Ara-C)	MA; MA	14–55	98/2/0/0	100	100	37	24	MMF+CsA+MTX	$\S^{d}$, α^{f}, β^{f}, ε^{cf}, μ^{f}, λ^{d}, ρ^{d}, π^{d}
Liu 2013b[47]	108	BU+CY BU+FLUD	MA; MA	12–52	0/91/0/9	85	87	NR	28	MMF+CsA+MTX	∞^{c}, $\S^{d}$, α^{c}, β^{ce}, ε^{d}, ψ^{d}
Kim 2010/ 2012[a42, 43]	83	CY+FLUD+ATG CY+ATG	Non-MA; Non-MA	15–60	71/25/1/2	NR	NR	At least 72	0	CsA+MTX+Steroids	∞^{f}, $\S^{f}$, β^{f}, μ^{f}, λ^{f}, ρ^{f}, ψ^{f}
Bornhauser 2011[33]	195	CY+TBI; FLUD+TBI	MA; Non-MA	18–60	11/89/0/0	60	>60 *	NR	38	MTX+CsA (ATG for UD grafts)	$\S^{d}$, α^{c}, μ^{f}, λ^{d}, ρ^{d}, ψ^{d}
Champlin 2007[a34]	130	CY CY+ATG	Non-MA; Non-MA	1–51	97/2/1/0	100	100	26	0	CsA+MTX	∞^{f}, α^{f}, β^{f}
Ringdén 1994/1999[49-51]	167	BU+CY; CY+TBI	MA; MA	1–55	100/0/0/0	100	100	NR	28	MTX+CsA	∞^{c}, $\S^{d}$, α^{c}, β^{e}, ε^{c}, μ^{c}, γ^{d}, λ^{d}, ρ^{d}
Devergie 1995[37]	120	CY+TBI; BU+CY	MA; MA	10–54	100/0/0/0	100	100	NR	0	CsA+MTX (+MoAB to 13%)	∞^{c}, α^{c}, ε^{c}, μ^{d}, λ^{d}
Clift 1994/ 1999[35, 36]	142	CY+TBI; BU+CY	MA; MA	6–55	100/0/0/0	100	100	25	0	MTX+CsA	∞^{c}, $\S^{d}$, α^{c}, β^{c}, ε^{cd}, μ^{d}, ϕ^{d}, ρ^{d}, ψ^{d}
Blume 1993[32]	122	ETP+TBI; BU+CY	MA; MA	2–48	100/0/0/0	100	100	NR	56	CsA+Steroids	∞^{c}, $\S^{d}$, α^{c}, β^{e}, ε^{c}
Gratwhol 1992/ 1993[b39, 40]	235	TLI+Standard CR; Standard CR	MA; MA	8–52	100/0/0/0	100	100	27	0	T-cell depletion (25%), CsA (17%), MTX+CsA (46%),	α^{f}

Table 4: Summary of characteristics of studies included in the systematic review of conditioning regimens for allogeneic HSCT

Author (year)	# of patients	Conditioning regimens compared	Intensity of Conditioning Regimens	Patient age range (years)	Donor source (%BM/PB/ UC/BM+PB)	% related donor	% full HLA match	% female donor / male recipient	% high risk	GVHD prophylaxis	Endpoints reported
										other (12%)	
Blaise 1992/ 2001[29, 30]	101	CY+TBI; BU+CY	MA; MA	24–40	100/0/0/0	100	100	NR	NR	CsA+MTX (+MoAB to 17%)	∞[c], §[d], α[c], β[e], ε[c], μ[c], ϕ[d], λ[d]
Helenglass 1988[41]	63	CY+TBI; MELPH+TBI	MA; MA	2–48	100/0/0/0	100	100	NR	NR	CsA+Steroids	∞[c], §[d], α[c], ε[d], μ[c], ψ[d]

Studies are listed in reverse chronologic order by publication date.

[a]Enrolled only aplastic anaemia patients and not included in NMA; summarized narratively in appendix

[b]Study could not be included in NMAs and has been summarized narratively

[c]Outcome included in NMAs in the main body of the report

[d]Outcome summarized narratively in the main body of the report

[e]Outcome included in NMAs in appendices

[f]Outcome summarized narratively in appendices

∞ = overall mortality; § = non-relapse mortality; α = acute GVHD; β = chronic GVHD; ε = relapse; μ = veno-occlusive disease or hepatic dysfunction; γ = bronchiolitis obliterans; ϕ = sepsis or bacterial infection; λ = pneumonia or interstitial pneumonitis; ρ = cytomegalovirus antigenemia or reactivation; θ = fungal infection; π = other infection; ψ = other harms

*60% matched siblings and 40% with 9 or 10/10 matched HLA alleles

Ara-C = cytosine arabinoside; ATG = anti-thymocyte globulin; BM = bone marrow; BU = busulfan; CR = conditioning regimen; CsA = cyclosporine A; CY = cyclophosphamide; ETP = etoposide; FLUD = fludarabine; HLA = human leukocyte antigens; HU = hydroxyurea; MA = myeloablative; Me-CCNU = semustine; MELPH = melphalan; MoAB = anti-P55 interleukin-2 receptor monoclonal antibody; MMF = mycophenolate mofetil; MTX = methotrexate; Non-MA = non-myeloablative; NMA = network meta-analysis; NR = not reported; PB = peripheral blood; TAC = tacrolimus; TBI = total body irradiation; THIO=thiotepa; TLI = total lymphoid irradiation; UC = umbilical cord blood; UD = unrelated donor

3.4. Overview of network geometry and interventions compared

Generally, the networks of the evidence available for each of the outcomes of interest had high proportions of single-study connections. **Table 5** presents the numbers of interventions, studies, and single-study connections available for each of the outcomes assessed. **Table 6** provides an overview of the distribution of treatment comparisons across included studies in reverse chronologic order.

8.5 Findings from Risk of Bias Assessment

Risk of bias assessment was possible for all 18 studies (100%). Adequate random sequence generation was clearly reported in 3 of the studies (33%), while only 1 study had a low risk of bias for allocation concealment (6%). The methods of random sequence generation and allocation concealment were not clearly reported for the remaining 15 studies (67%) and 17 studies (94%), respectively, and these studies were assigned an unclear risk of bias for these domains. Patients and investigators could not be blinded from regimen allocation in any of the studies (0%) due to the nature of the interventions being evaluated. Independent and blinded outcome assessment was also not reported in any of the studies (0%), with 2 of the studies being assigned a high risk of bias given lack of blinding, and 16 studies (89%) being assigned an unclear risk of bias as they did not report sufficient information regarding this domain. A total of 15 studies (83%) used intention-to-treat (ITT) or modified ITT analysis for the main outcomes, and therefore had low risk of attrition bias. One study (6%) had a high risk of attrition bias and 2 (11%) had an unclear risk of bias. Eight studies (44%) provided reference to a study protocol and had low risk of bias for selective outcome reporting, while this information was either not reported in 9 (50%), or suggested high risk in 1 studies (6%). Other potential sources of bias deemed important by the reviewers included imbalances in baseline patient characteristics, including co-interventions administered and other potentially important factors. Data for baseline variables and co-intervention were poorly reported in 5 studies (28%), preventing the reviewers from objectively assessing these potential other biases, rendering the risk for this domain unclear. Totals of 11 (61%), and 2 (11%) of studies had low and high risk of other biases, respectively. A complete listing of risk of bias assessments for all included studies is presented in **Appendix 3**. A summary of the distribution of risk of bias across studies is provided in **Figure 2**.

.

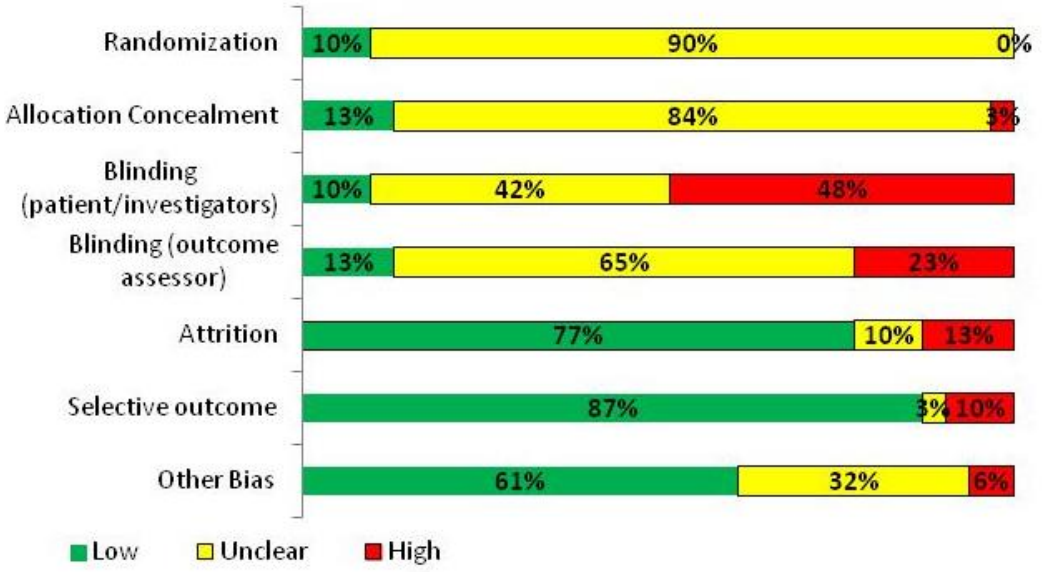
Table 5: Summary of interventions compared, number of studies, and number of single-study connections by outcomes analysed by NMA for conditioning regimens					
Clinical Outcome	**Number of interventions**	**Number of studies reporting outcome**	**Number of single-study connections**	**Number of head-to-head connections**	**Total number of possible connections**
Overall mortality at 100 days, primary network	5	8	2	4	10
Overall mortality at 100 days, secondary network	4	3	3	3	6
Overall mortality at 1 year, primary network	5	9	2	4	10
Overall mortality at 1 year, secondary network	4	3	3	3	6
Overall mortality at 2 years	5	8	2	4	10
Overall mortality at 4 years	5	8	2	4	10
Overall mortality at 5 years, primary network	5	7	2	4	10
Overall mortality at 5 years, secondary network	4	3	3	3	6
Acute GVHD	9	11	7	8	36
Chronic GVHD, extensive in evaluable patients, primary network	3	4	0	2	4
Chronic GVHD, extensive in evaluable patients, secondary network	3	2	2	2	3
Chronic GVHD, overall GVHD in evaluable patients	4	4	2	3	6
Relapse 2–3 years	4	7	1	3	6
Veno-occlusive disease or hepatic dysfunction, ≥100 days	4	4	2	3	6

Table 6: Interventions evaluated in trials of conditioning regimens for allogeneic HSCT, ordered by year of publication

Author	Year	CY+TBI	MELPH+TBI	BU +CY	ETP +TBI	FLUD +TBI	Modified BU+CY	Modified BU+FLUD	BU+FLUD	TBI	ATG+ FLUD+BU	ALZ+ ELPH+ FLUD	ATG+ THIO+ MELPH+ CY	ATG+TLI
Baron[27, 28]	2014/ 2015					X								X
Rambaldi[38, 48]	2014/ 2015			X					X					
Algarotti[26, 52]	2013											X	X	
Blaise[31]	2013					X					X			
Kornblit[44]	2013					X				X				
Lee[45]	2013			X					X					
Liu[46]	2013						X	X						
Liu[47]	2013			X					X					
Bornhauser[33]	2011	X				X								
Devergie[37]	1995	X		X										
Ringdén[49-51]	1994/ 1999	X		X										
Clift[35, 36]	1994/ 1999	X		X										
Blume[32]	1993			X	X									
Blaise[29, 30]	1992/ 2001	X		X										
Helenglass[41]	1988	X	X											

ATG = anti-thymocyte globulin; BU = busulfan; CY = cyclosporine; ETP = etoposide; FLUD = fludarabine; MELPH = melphalan; TBI = total body irradiation; THIO = thiotepa; TLI = total lymphoid irradiation.

Figure 2: Summary of Risk of Bias Assessments of Included Trials, Conditioining Regimens (n = 18 studies)



3.5. Findings: Overall mortality

Twelve studies[27, 29, 31, 32, 35, 37, 41, 44, 45, 47, 48, 51] evaluating 1,512 patients had data available for at least one of the overall mortality follow-up periods assessed. Nine conditioning regimens were evaluated in these studies; however, these interventions were distributed in two disjoint networks of 5 and 4 interventions each (**Figure 3**). Consultation with content experts determined that the most clinically relevant network was the larger of the two, containing CY+TBI, BU+CY, ETP+TBI, MELPH+TBI, and BU+FLUD. This network was thus chosen to be the focus of analyses, with the smaller network, containing FLUD+TBI, ATB+TLI, TBI, and ATG+FLUD+BU, being summarized in the report's supplement.

Figure 3: Network diagrams for the NMA of overall mortality

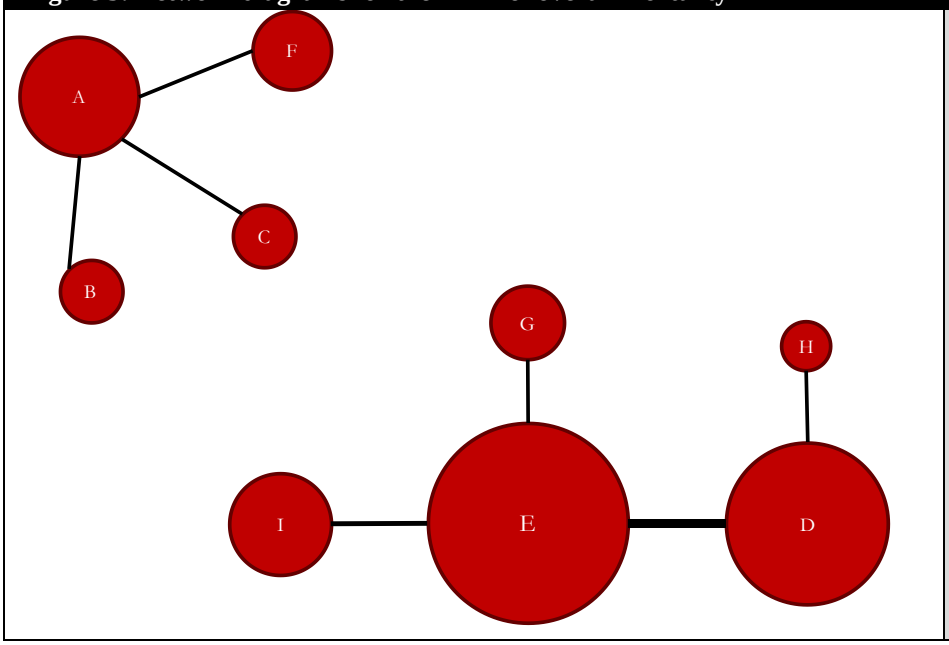


Figure legend
A = FLUD+TBI
B = ATG+TLI
C = TBI
D = CY+TBI
E = BU+CY
F = ATG+FLUD+BU
G = ETP+TBI
H = MELPH+TBI
I = BU+FLUD

ATG = anti-thymocyte globulin
BU = busulfan
CY = cyclophosphamide
ETP = etoposide
FLUD = fludarabine
MELPH = melphalan
TBI = total body irradiation
TLI = total lymphoid irradiation

3.5.1. Results from traditional pairwise meta-analyses

For each of the reported follow-up times for overall mortality, **Table 7** present summaries of pairwise estimates that were derived from direct evidence (i.e., the head-to-head trials), summarized alongside the numbers of available trials and patients for each direct comparison, as well as the related summary estimate derived from network meta-analysis. Eight studies were included in NMAs for 100-day and 2- and 4-year mortality[29, 32, 35, 37, 41, 45, 47, 51], while nine studies were included in the NMA for 1-year mortality[29, 32, 35, 37, 41, 45, 47, 48, 51], and seven studies for 5-year mortality[29, 32, 35, 37, 41, 47, 48].

The direct comparisons with the largest numbers of studies available were CY+TBI vs. BU+CY (4 RCTs, n = 530) and BU+FLUD vs. BU+CY (2 RCTs, n = 234). None of the comparisons demonstrated statistically significant differences in risk of mortality on pairwise meta-analysis. One of the comparisons (MELPH+TBI vs. CY+TBI)[41] had no mortality events in either intervention group, making a direct estimate from pairwise meta-analysis non-estimable. Estimates from traditional pairwise meta-analyses compared well with estimates derived from the FE network meta-analysis.

The direct comparisons with the largest numbers of studies available were CY+TBI vs. BU+CY and BU+FLUD vs. BU+CY. The numbers of studies included and the number of patients included varied slightly by time point as not all studies reported data up to 5 years of follow-up and for one comparison

(ETP+TBI vs. BU+CY) the authors reported data for only the "good risk" subgroup of patients after 3 years of follow-up[32].

Table 7: Summary of results from pairwise meta-analysis and NMA, overall mortality at multiple follow-up times

Comparison		# of Trials (patients)	Heterogeneity (I^2)	Direct Estimate OR (95% CI)	FE NMA Estimate OR (95% CrI)
				*Values <1 favor comparator	
Comparator	**Reference**				
100-day overall mortality					
BU+FLUD	BU+CY	2 (234)	0.000	0.51 (0.15–1.72)	0.41 (0.10–1.37)
ETP+TBI	BU+CY	1 (122)	—	0.58 (0.23–1.47)	0.57 (0.22–1.42)
CY+TBI	BU+CY	4 (530)	0.000	0.79 (0.45–1.36)	0.81 (0.47–1.40)
MELPH+TBI	CY+TBI	1 (63)	—	Not estimable*	0.76 (<0.01–288.90)
1-year overall mortality					
CY+TBI	MELPH+TBI	1 (63)	—	0.82 (0.30–2.29)	0.82 (0.29–2.35)
CY+TBI	BU+CY	4 (530)	0.000	**0.59 (0.39–0.90)**	**0.61 (0.40–0.93)**
BU+FLUD	BU+CY	3 (479)	72.457	0.90 (0.58–1.40)	0.90 (0.58–1.40)
BU+CY[a]	ETP+TBI[a]	1 (122)	—	0.71 (0.35–1.47)	0.70 (0.34–1.47)
2-year overall mortality					
CY+TBI	MELPH+TBI	1 (63)	—	1.00 (0.37–2.73)	1.00 (0.36–2.81)
CY+TBI	BU+CY	4 (530)	42.008	**0.52 (0.35–0.77)**	**0.51 (0.34–0.75)**
BU+CY[a]	ETP+TBI[a]	1 (122)	—	1.00 (0.47–2.15)	1.00 (0.46–2.14)
BU+CY	BU+FLUD	2 (234)	85.761	0.62 (0.35–1.13)	0.62 (0.34–1.11)
4-year overall mortality					
MELPH+TBI	CY+TBI	1 (63)	—	1.00 (0.37–2.73)	1.01 (0.36–2.83)
CY+TBI	BU+CY	4 (530)	25.640	**0.60 (0.41–0.87)**	**0.62 (0.43–0.90)**
ETP+TBI[b]	BU+CY[b]	1 (48)	—	0.38 (0.12–1.26)	0.37 (0.11–1.21)
BU+CY	BU+FLUD	2 (234)	75.275	0.81 (0.46–1.43)	0.81 (0.46–1.40)
5-year overall mortality					
MELPH+TBI	CY+TBI	1 (63)	—	0.72 (0.26–1.95)	0.70 (0.25–1.96)
ETP+TBI[b]	BU+CY[b]	1 (48)	—	0.38 (0.12–1.26)	0.37 (0.11–1.22)
CY+TBI	BU+CY	3 (355)	42.788	0.66 (0.42–1.03)	0.65 (0.41–1.02)
BU+FLUD	BU+CY	2 (353)	0.000	0.85 (0.55–1.33)	0.85 (0.54–1.33)

*There were zero deaths in both intervention groups, making the direct estimate for MELPH+TBI vs. CY+TBI not estimable. [a]Both "good risk" and "poor risk" patients included; [b]Only "good risk" patients included

3.5.2. Results from network meta-analysis

For all mortality follow-up periods, 6 of the 10 possible pairwise comparisons were informed only by indirect evidence and half of the comparisons with direct evidence were informed by single studies with limited numbers of patients (**Table 5**). Model fit statistics for both FE and RE NMAs are summarized in **Table 8**. DIC values suggested some difference in fit favouring the RE model for 2-year mortality, however, a FE model was preferred due to the number of single-study connections and concerns regarding inflated estimates of the between-study variance parameter (RE findings are reported in the report's supplement).

Table 8: Model fit for 100-day, 1-, 2-, 4-, and 5-year OM networks in a NMA

Follow-up time	# of comparisons informed only by indirect evidence of total possible comparisons	# of single-study direct comparisons	Number of data points	FE total deviance residual	RE total deviance residual	FE DIC	RE DIC
100-day	6 of 10	2 of 4	16	16.21	16.32	77.899	79.384
1-year	6 of 10	2 of 4	18	23.35	18.69	112.579	111.420
2-year	6 of 10	2 of 4	16	24.64	16.63	104.993	100.438
4-year	6 of 10	2 of 4	16	20.01	16.53	99.921	99.340
5-year	6 of 10	2 of 4	14	15.52	14.11	86.411	87.011

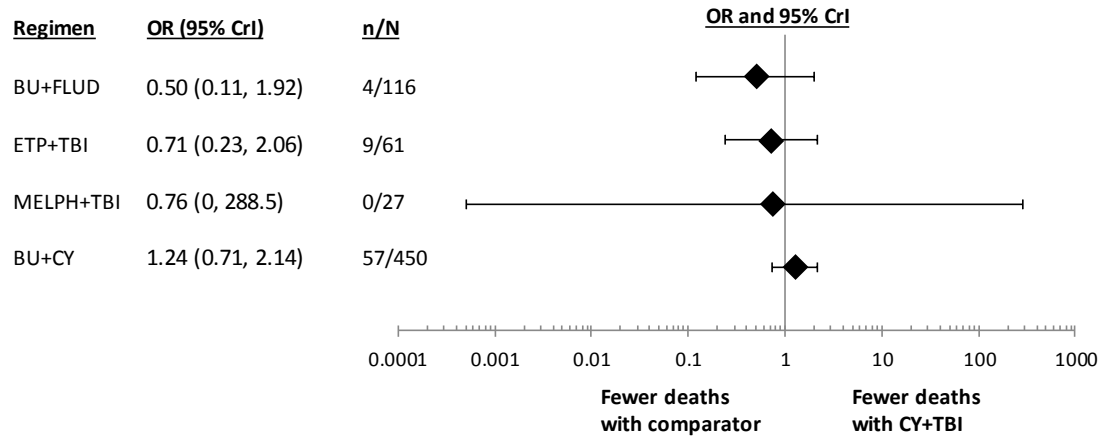
DIC = deviance information criterion; FE = fixed effects mode; RE = random effects model

3.5.2.1. Comparisons versus CY+TBI

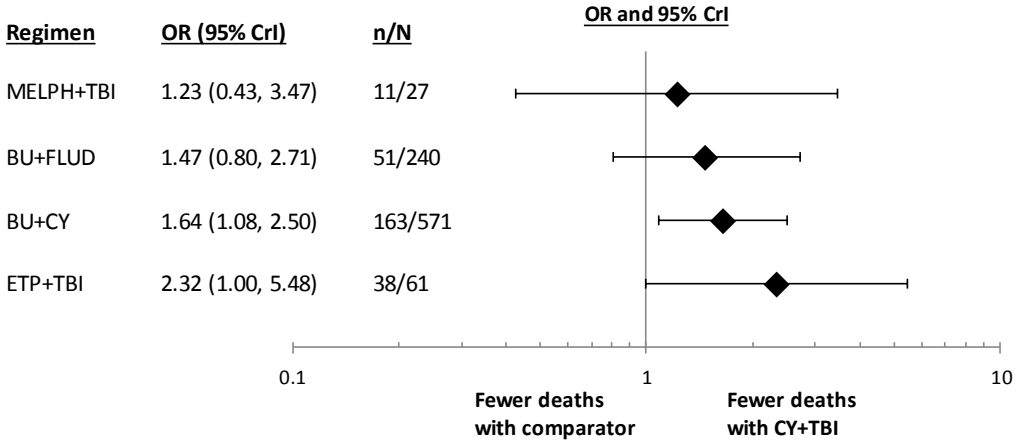
Figure 4a (panels A-E) presents forest plots summarizing comparisons between all conditioning regimens in the evidence network to the chosen reference regimen, CY+TBI; findings from all time points analysed are shown. Interventions associated with credible intervals that include 1 are not significantly different from CY+TBI in reducing overall mortality at the selected follow-up time. Based upon results from the FE NMA analysis, CY+TBI was associated with a reduced risk of mortality at (a) 1 year when compared to BU+CY and ETP+TBI; (b) 2 years when compared to BU+CY and BU+FLUD; and (c) 4 years when compared to BU+CY and BU+FLUD. No other conditioning regimens were significantly different from CY+TBI with respect to overall mortality at any of the analysed follow-up times. Findings from the corresponding RE NMAs provided in the report's supplement were all associated with wider 95% credible intervals wherein none of the aforementioned significant differences remained; these findings should thus be interpreted cautiously.

Figure 4a: Forest plots of comparisons versus CY+TBI, Overall mortality, multiple follow-up times

A: 100-day overall mortality



B: 1-year overall mortality



C: 2-year overall mortality

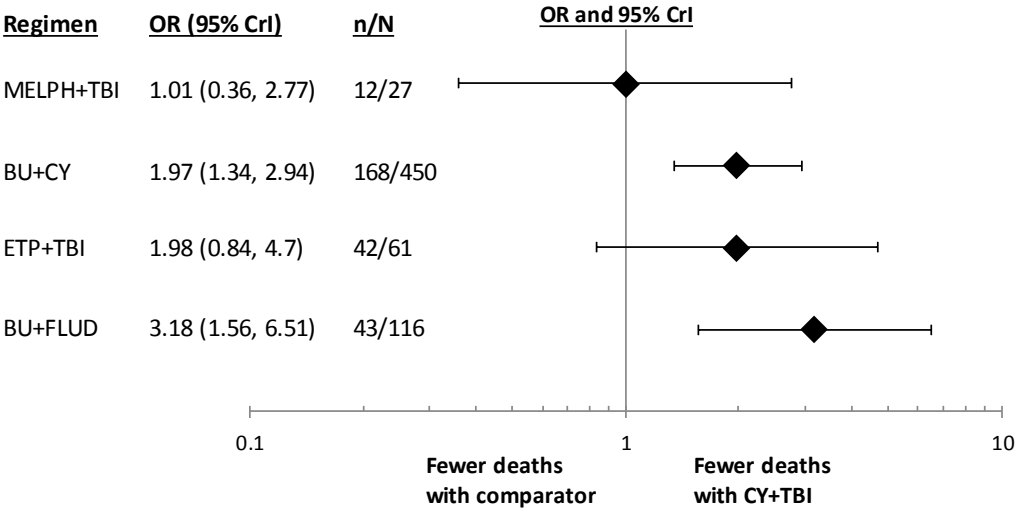
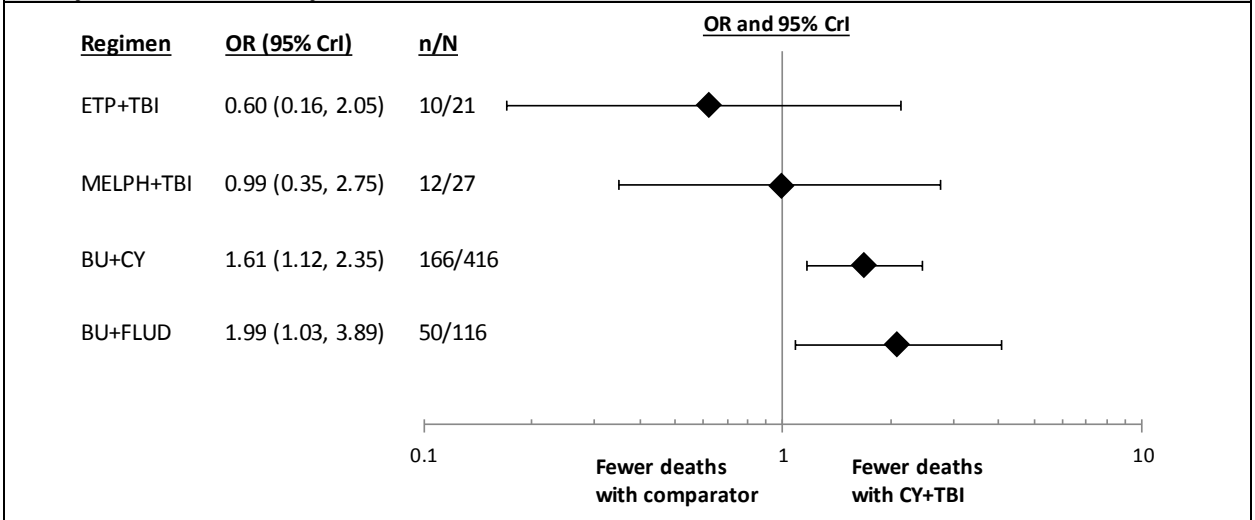
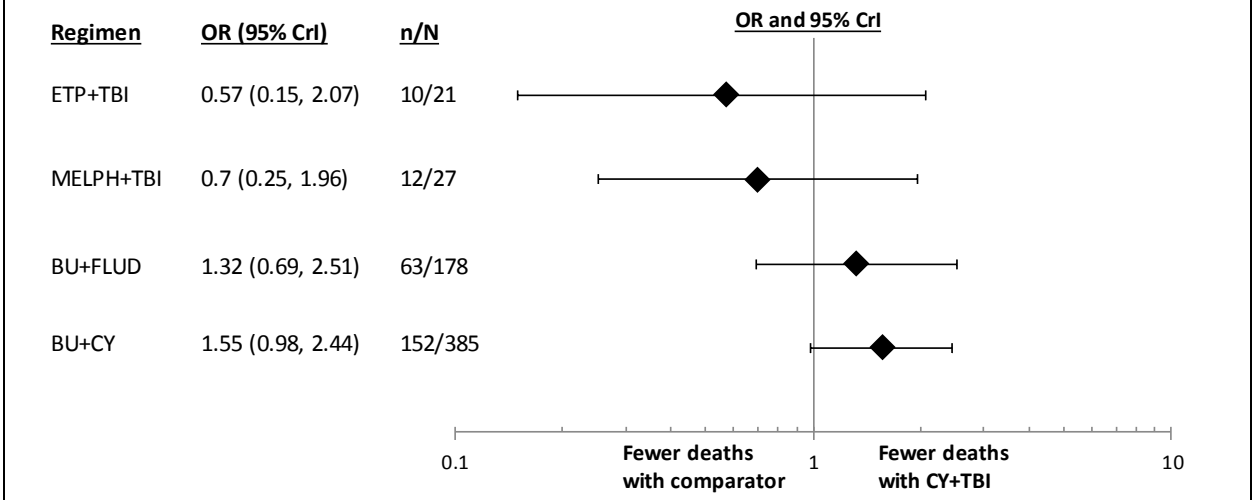


Figure 4a: Forest plots of comparisons versus CY+TBI, Overall mortality, multiple follow-up times

D: 4-year overall mortality



E: 5-year overall mortality



Odds ratios from network meta-analysis comparing each regimen to the reference treatment of CY+TBI are presented along with the number of events and patients per regimen. Odds ratios <1 suggest more risk of mortality with CY+TBI while odds ratios >1 suggest less risk of mortality with CY+TBI. All results are generated from a fixed effects NMA; results from random effects analyses are provided in the report's appendices.

3.5.2.2. Comparisons between all conditioning regimens

Figure 4b presents league tables of the estimates for all 10 pairwise comparisons generated by NMA for each of the 5 follow-up times. Some of the comparisons included in the league tables should be interpreted carefully, specifically those which are not informed by head-to-head trials or where the comparison was based on compound indirect evidence. There are several specific aspects of the data that should be taken into consideration while interpreting the data:

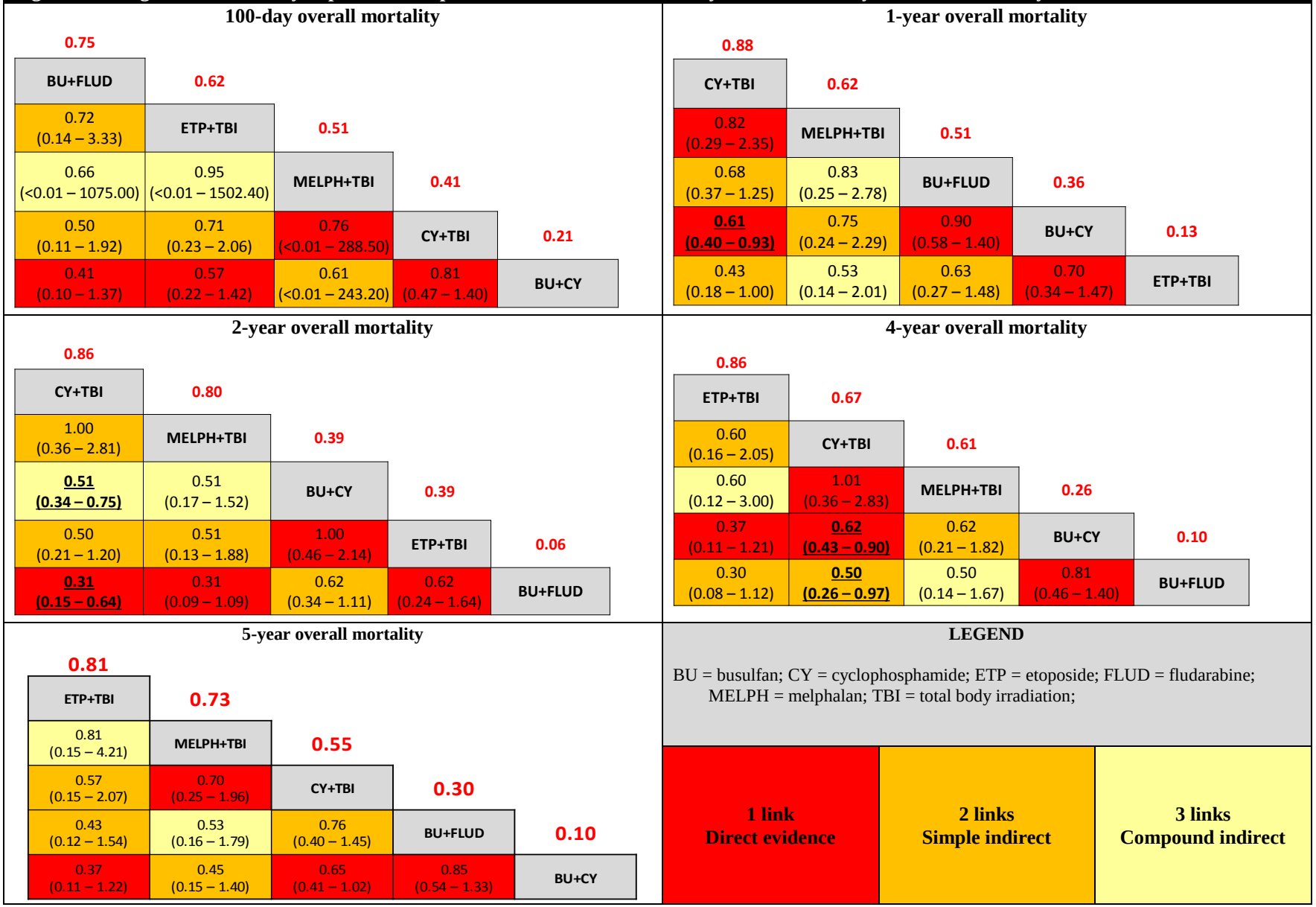
1) There were zero mortality events prior to 100 days for the regimen MELPH+TBI, meaning that confidence intervals for all comparisons with this regimen are extremely wide, reflecting the high level of uncertainty in the comparison.

2) The single study[32] evaluating ETP+TBI included patients of all disease risk levels up to 2 years of follow-up; however, data were only available for "good risk" patients for NMAs performed using data available at 4 and 5 year follow-up. Consequently, the ranking of ETP+TBI at 4 and 5 year follow-up is likely artificially higher relative to the other interventions in the network. When the study evaluating ETP+TBI was removed from the network in a secondary analysis, the effect sizes of the remaining comparisons remained similar to the full model, and there were no changes in significance or ranking of interventions.

3) Different studies and patient samples contributed to the data at different follow-up times. Therefore, evaluating trends of relative rankings of the interventions over time should be done with caution.

For 100-day mortality, all summary odds ratios included the null value of 1 and many were associated with very wide credible intervals, providing no evidence of the presence of important differences between regimens. Inspection of 1-year data did not identify any additional significant differences beyond the advantages for CY + TBI noted above. The same was true for mortality findings at 2 years. No additional significant differences between regimens were noted from FE NMAs at either 4 years or 5 years of follow-up; while ETP+TBI was associated with the most favourable SUCRA values for both of these follow-up times, these results are limited by the above noted issue regarding the limited access to patient data in regard to their risk level. League tables of findings from the corresponding RE NMAs are provided in the report's supplement, with consistently greater uncertainty in findings which also were associated with no evidence of important differences between regimens.

Figure 2b: League table summary of pairwise comparisons from FE NMA, 100-day, 1-, 2-, 4-, and 5-year overall mortality



3.6. Findings: Non-relapse Mortality

Network meta-analysis was considered infeasible for the non-relapse mortality (NRM) outcome due to the competing risk of relapse, which changed the number of patients at risk over the follow-up period. Overall patient time at risk for NRM was not reported in most included studies. Where patient time at risk was reported, the duration of follow-up varied, preventing an incidence-rate-based analysis. Consequently, a narrative summary of NRM was prepared.

Fourteen studies[26, 27, 29, 31-33, 35, 41, 44-48, 51] reported results for NRM at various follow-up times. The findings from these studies are presented in **Table 9.**

Several studies reported statistically significant differences between treatment arms[31, 33, 44, 48, 51]; BU+CY was compared to BU+FLUD in 3 studies[45, 47, 48]. At 1, 2, and 5 years of follow-up, the cumulative incidence of NRM was significantly higher in patients receiving BU+CY compared to those receiving BU+FLUD[48]. However, for the same regimens, no significant difference in NRM was identified at 2 or 5 years in other studies[45, 47].

CY+TBI was compared to BU+CY in 3 studies[29, 35, 51]. After both early (0.1–4.2 years) and late (1.7–9.3 years) follow-up durations, BU+CY demonstrated statistically significantly higher NRM than CY+TBI[51] ($p = 0.006$ and $p = 0.01$, respectively). These differences became increasingly significant in the subgroup of high risk patients with advanced disease ($p < 0.002$ and $p = 0.004$ for 3- and 7-year K-M estimates, respectively). When a multivariable survival model was fit to these data to adjust for the effects of disease status (early vs. advanced) and donor age (< vs. > 30 years), the difference in the hazard of NRM for BU+CY compared to CY+TBI was not significant ($p = 0.05$). In other studies that compared the same regimens, no significant difference in NRM was found at 2, 3, 9 and 10 years of follow-up[29, 35].

FLUD+TBI was compared to 3 other regimens in studies[27, 31, 33, 44]. Compared to FLUD+TBI, NRM was significantly increased with ATG+BU+FLUD at 5 years of follow-up, but not at 1 year[31]. In another comparison with FLUD+TBI, a near significant increase in NRM was found with the use of CY+TBI at 1 year of follow-up[33]. This difference reached statistical significance in the patient subgroup >40 years of age ($p = 0.01$). There were no significant differences in NRM between FLUD+TBI and ATG+TLI at 4 years ($p = 0.5$)[27, 28] or TBI alone at 3.5 years ($p = 0.35$)[44].

Table 9: Overview, incidence and risk of non-relapse mortality in the evaluation of conditioning regimens

Author	Follow-up Median (range) unless otherwise specified	Reporting format	Treatment regimen	Findings by treatment arm	Study conclusions
Liu (2013)[46]	Maximum 3.3 years	2-year cumulative incidence	mBU+FLUD	23% of 52	NSD ($p = 0.169$)
			mBU+CY	13.2% of 53	
Blaise (2013)[31]	5 years	5-year Kaplan-Meier estimate	ATG+BU+FLUD	38% (95% CI: 27–48) of 69	**Significant difference ($p = 0.027$)**
			FLUD+TBI	22% (95% CI: 12–32) of 70	
Blaise (2013)[31]	1 year	1-year Kaplan-Meier estimate	ATG+BU+FLUD	17% (95% CI: 9–26) of 69	NSD
			FLUD+TBI	11% (95% CI: 4–18) of 70	

Table 9: Overview, incidence and risk of non-relapse mortality in the evaluation of conditioning regimens					
Author	**Follow-up Median (range) unless otherwise specified**	**Reporting format**	**Treatment regimen**	**Findings by treatment arm**	**Study conclusions**
Blaise (2013)[31]	4.5 (2.2–7.3) years	Frequency	ATG+BU+FLUD	23/69 (33%)	—
			FLUD+TBI	11/70 (16%)	
Blume (1993)[32]	2.6 (0.8–4.8) years	Frequency	ETP+TBI	17/55 (31%)	NR
			BU+CY	21/59 (36%)	
Bornhauser (2011)[33]	1 year	1-year cumulative incidence	CY+TBI	17% (95% CI: 9–24%) of 90 per protocol analysis (ITT n = 96, NRM not reported)	**Per protocol analysis: significant difference (p = 0.048)** **ITT analysis: NSD (p = 0.06)**
			FLUD+TBI	8% (95% CI: 3–14) of 94 per protocol analysis (ITT n = 99, NRM not reported)	
Clift (1994)[35, 36]	3 years	3-year Kaplan-Meier estimate	CY+TBI	24% of 69	NSD
			BU+CY	18% of 73	
Clift (1994)[35, 36]	9 years	9-year cumulative incidence	CY+TBI	25% of 69	NSD
			BU+CY	20% of 73	
Helenglass (1988)[41]	Range: 1–5.5 years	Frequency	CY+TBI	10/36 (28%)	NR
			MELPH+TBI	11/27 (41%)	
Lee (2013)[45]	2 years	Kaplan-Meier estimate	BU+CY	18.7% of 64	NSD (p = 0.235)
			BU+FLUD	34.3% of 62	
Lee (2013)[45]	2.2 (0.3–4.7) years	Frequency	BU+CY	11/64 (17%)	—
			BU+FLUD	15/62 (24%)	
Lee (2013)[45]	2.2 (0.3–4.7) years	Univariable survival model estimate	BU+CY*	—	NSD: HR = 1.603; 95% CI: 0.731–3.517 (p = 0.239)
			BU+FLUD	—	
Lee (2013)[45]	2.2 (0.3–4.7) years	Multivariable survival model	BU+CY*	—	NSD: HR = 2.245; 95% CI: 0.952–5.296 (p = 0.065)
			BU+FLUD	—	
Liu (2013)[47]	1.7 (0.01–5.9) years	Frequency	BU+CY	8/54 (15%)	NSD (p = 0.112)
			BU+FLUD	3/54 (6%)	
Liu (2013)[47]	5 years	5-year cumulative incidence	BU+CY	18.8% (95% CI: 11.9–25.7) of 54	NSD (p = 0.104)
			BU+FLUD	9.9% (95% CI: 3.6–16.2) of 54	
Baron (2015)[27, 28]	4 years	4-year cumulative incidence	FLUD+TBI	24% of 49	NSD (p = 0.5)
			ATG+TLI	13% of 45	
Rambaldi (2014)[38, 48]	1 year	1-year cumulative incidence	BU+CY	17.4% of 121	**Significant difference (p = 0.02)**
			BU+FLUD	7.3% of 124	
Rambaldi	2 years	2-year	BU+CY	18.2% of 121	**Significant**

Table 9: Overview, incidence and risk of non-relapse mortality in the evaluation of conditioning regimens					
Author	**Follow-up Median (range) unless otherwise specified**	**Reporting format**	**Treatment regimen**	**Findings by treatment arm**	**Study conclusions**
(2014)[38, 48]		cumulative incidence	BU+FLUD	8.9% of 124	**difference (p = 0.03)**
Rambaldi (2014)[38, 48]	5 years	5-year cumulative incidence	BU+CY	19.0% of 121	**Significant difference (p = 0.03)**
			BU+FLUD	9.7% of 124	
Algarotti (2013)[26, 52]	2 years	2-year cumulative incidence	ALZ+MELPH+FLUD	35% of 58	NSD
			ATG+THIO+MELPH+CY	36% of 54	
Blaise (1992)[29, 30]	At 1.5 years	Frequency	CY+TBI	4/49 (8%)	NSD (p = 0.06)
			BU+CY	11/51 (22%)	
Blaise (1992)[29, 30]	Mean = 1.9 ± 0.9 years (0.5–3.7 years)	Kaplan-Meier estimate	CY+TBI	8% (95% CI: 5–11) of 49	NSD (p = 0.06)
			BU+CY	27% (95% CI: 20–34) of 51	
Blaise (1992)[29, 30]	10.8 (9.5–12.7) years	Frequency	CY+TBI	9/49 (18%)	NSD
			BU+CY	14/51 (28%)	
Ringdén (1999)[49-51]	Range: 0.1–4.2 years	Cumulative incidence	CY+TBI	9% of 79	**Significant difference (p = 0.006)**
			BU+CY	28% of 88	
Ringdén (1999)[49-51]	Range: 0.1–4.2 years	Multivariable survival model[a]	CY+TBI*	—	NSD: HR = 2.40; 95% CI: 0.98–5.85 (p = 0.05)
			BU+CY	—	
Ringdén (1999)[49-51]	CY+TBI: 6.9 (2.3–8.7) years BU+CY: 7.3 (1.7–9.3) years	Frequency	CY+TBI	11/79 (14%)	—
			BU+CY	27/88 (31%)	
Ringdén (1999)[49-51]	CY+TBI: 6.9 (2.3–8.7) yrs BU+CY: 7.3 (1.7–9.3) ys	7-year cumulative incidence	CY+TBI	14% of 79	**Significant difference (p = 0.01)**
			BU+CY	34% of 88	
Kornblit (2013)[44]	4.6 (0.6–7) years	3-year cumulative incidence	FLUD+TBI	14% of 41	NSD (p = 0.35)
			TBI	17% of 44	
Kornblit (2013)[44]	4.6 (0.6–7) years	Frequency and univariable survival estimate	FLUD+TBI	3/41 (7%)	NSD: HR = 0.67; 95% CI: 0.1–3.0 (p = 0.59)
			TBI*	4/44 (9%)	

Sample sizes of cumulative incidence data and denominators of frequency data are the original number of patients randomized (i.e., the ITT sample), unless otherwise noted.

*Denotes referent group for hazard ratio

[a]Hazard ratio adjusted for the effects of disease status (early vs. advanced) and donor age (< or > 30 years)

ATG = anti-thymocyte globulin; BU = busulfan; CY = cyclophosphamide; FLUD = fludarabine; HR = hazard ratio; mBU+CY = BU+CY+ hydroxyurea+semustine+cytosine arabinoside; mBU+FLUD = BU+FLUD+hydroxyurea+semustine+cytosine arabinoside; NR = not reported; NSD = no significant difference; TBI = total body irradiation

3.7. Findings: Acute GVHD

Fourteen studies[26, 27, 29, 31-33, 35, 37, 40, 41, 44-46, 51] that enrolled a total of 1,773 patients had data available for the occurrence of aGVHD. Sixteen conditioning regimens were evaluated in these studies; however, six of these interventions were disjoint from the main network—there were 10 interventions in the main network and three separate pairs of interventions. One pair of disjoint interventions compared standard conditioning regimens to TLI + standard conditioning regimens; however, "standard conditioning regimens" were not consistent within the study and the data could not be included in analyses (pair not shown in **Figure 5** below). Thus the main network, evaluating 11 studies[27, 29, 31-33, 35, 37, 41, 44, 45, 51], was slected as the focus of the analyses below, and the three studies[26, 40, 46] making up the disjoint pairs have been summarized narratively in the report's appendices.

Figure 5: Network diagrams for the NMA of acute GVHD

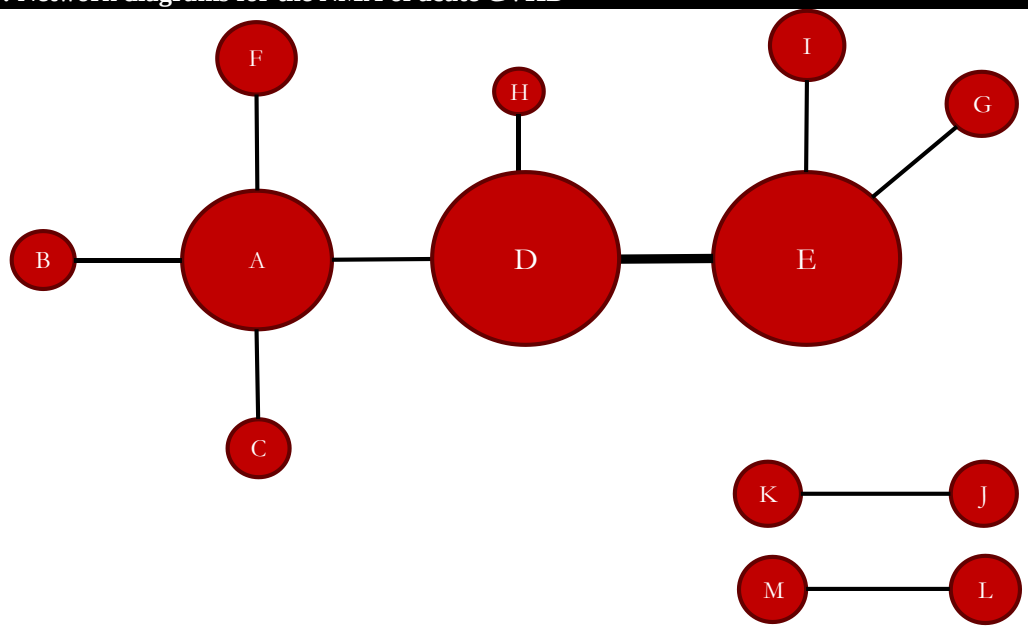


Figure legend

A = FLUD+TBI
B = ATG+TLI
C = TBI
D = CY+TBI
E = BU+CY
F = ATG+FLUD+BU
G = ETP+TBI
H = MELPH+TBI
I = BU+FLUD
J = mBU+FLUD
K = mBU+CY
L = ALZ+MELPH+FLUD
M = ATG+THIO+MELPH+CY

ALZ = alemtuzumab; ATG = anti-thymocyte globulin; BU = busulfan; CY = cyclophosphamide; ETP = etoposide; FLUD = fludarabine; mBU+CY = BU+CY+ hydroxyurea+semustine+cytosine arabinoside; mBU+FLUD = BU+FLUD+ hydroxyurea+semustine+cytosine arabinoside; MELPH = melphalan; TBI = total body irradiation; THIO = thiotepa; TLI = total lymphoid irradiation

3.7.1. Results from traditional pairwise meta-analyses

Table 10 presents summaries of pairwise estimates that were derived from direct evidence (i.e., the head-to-head trials), alongside the numbers of available trials and patients for each direct comparison, as well as the related summary estimate derived from the network meta-analysis.

The direct comparison with the largest number of studies available was BU+CY vs. CY+TBI (4 studies, 522 patients). All other comparisons were informed by single studies. Study data suggested a significantly reduced risk of acute GVHD with FLUD+TBI compared to ATG+BU+FLUD. The interventions in none of the other comparisons were significantly different from each other. Estimates from traditional pairwise meta-analyses compared well with estimates derived from the FE NMA analysis.

Table 10: Summary of results from meta-analysis and NMA, Acute GVHD

Comparison		# of Trials (patients)	Heterogeneity (I^2)	Direct Estimate OR (95% CrI)	FE NMA Estimate OR (95% CI)
				*Values <1 favor comparator	
Comparator	**Reference**				
TBI	FLUD+TBI	1 (85)	—	0.54 (0.22–1.31)	0.54 (0.22–1.29)
ATG+TLI	FLUD+TBI	1 (94)	—	0.70 (0.18–2.66)	0.67 (0.16–2.62)
BU+FLUD	BU+CY	1 (126)	—	0.69 (0.28–1.69)	0.67 (0.26–1.68)
FLUD+TBI	CY+TBI	1 (184)	—	0.62 (0.30–1.30)	0.63 (0.29–1.34)
FLUD+TBI	ATG+BU+FLUD	1 (139)	—	**0.44 (0.22–0.88)**	**0.43 (0.21–0.87)**
ETP+TBI	BU+CY	1 (114)	—	0.79 (0.31–1.98)	0.79 (0.30–2.00)
BU+CY	CY+TBI	4 (522)	0.000	0.83 (0.57–1.19)	0.83 (0.57–1.19)
CY+TBI	MELPH+TBI	1 (63)	—	0.40 (0.14–1.12)	0.39 (0.13–1.11)

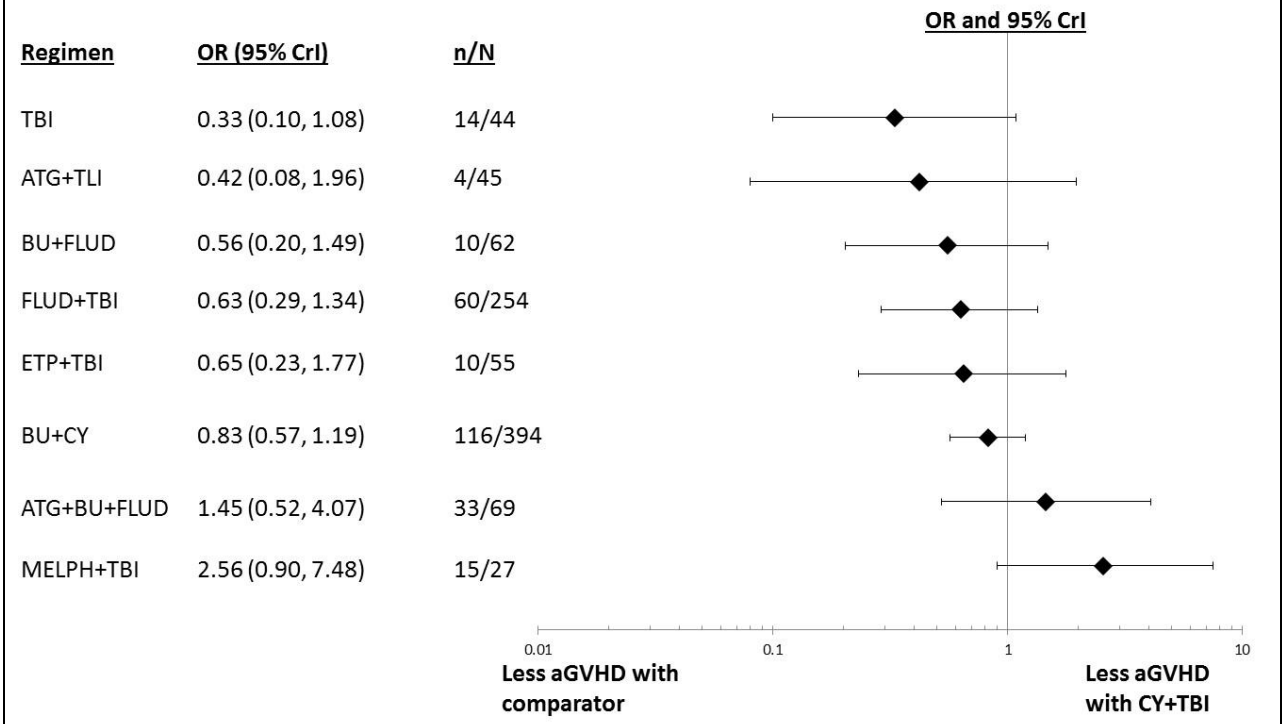
3.7.2. Results from network meta-analysis

Most comparisons (i.e., 28 of 36 possible) were informed only by indirect evidence and all but one of the comparisons with direct evidence were informed by single studies, often with limited numbers of patients. Model fit statistics for the NMA indicated that both the fixed-effects (FE) and random-effects (RE) models had an adequate fit. Posterior residual deviance values for the FE and RE models of 21.8 and 21.6, respectively, were obtained, both of which demonstrated a desirable fit, given the 22 data points in the model. DIC values (134.020 and 135.412, respectively) suggested similar fit; a FE model was preferred due to the number of single-study connections in the network (as for overall mortality, findings from RE NMAs are provided in the report's supplement).

3.7.2.1. Comparisons versus CY+TBI

Figure 6 presents a forest plot summarizing comparisons of all conditioning regimens in the evidence network to the chosen reference therapy, CY+TBI. All regimens were associated with credible intervals that included 1 and, thus, were not significantly different from CY+TBI in preventing aGVHD.

Figure 6: Forest plot of comparisons from NMA versus CY+TBI—Acute GVHD

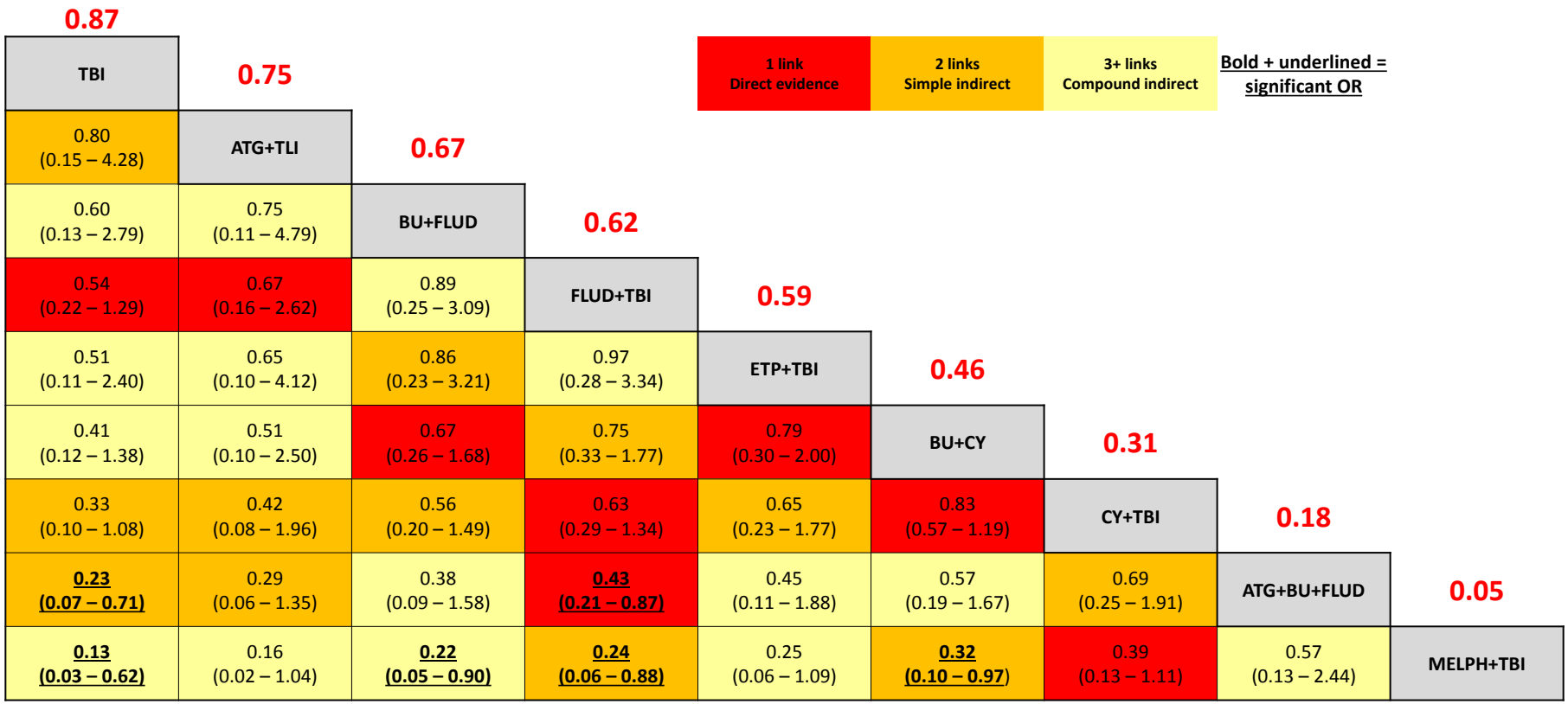


Odds ratios from network meta-analysis comparing each regimen to the reference regimen of CY+TBI are presented along with the number of events and patients per regimen. Odds ratios <1 suggest more risk of acute GVHD with CY+TBI while odds ratios >1 suggest less risk of acute GVHD with CY+TBI.

3.7.2.2. Comparisons between all conditioning regimens

Review of the league table summary of results shown in **Figure 7** from FE NMA a significantly reduced risk of aGVHD associated with TBI in comparison to both ATG+BU+FLUD and MELPH+TBI. FLUD+TBI was associated with a statistically significantly reduced risk compared to both ATG+BU+FLUD as well as MELPH+TBI. BY+CY was also shown to be associated with a significantly reduced risk compared to MELPH+TBI. Inspection of all other pairwise comparisons found that all 95% credible intervals included the null value of 1. In the corresponding RE NMA, all 95% credible intervals were notably wider and no pairwise comparisons were associated with statistically significant differences (see the review's supplement for numeric details). It should be noted that the prophylactic GVHD regimens used across studies with available outcome data varied considerably, and this variation is also likely to have had influence upon the findings observed.

Figure 7: League table summary of pairwise comparisons from FE NMA, Odds Ratios with 95% CrI, Acute GVHD



0.87								
TBI	0.75							
0.80 (0.15 – 4.28)	ATG+TLI	0.67						
0.60 (0.13 – 2.79)	0.75 (0.11 – 4.79)	BU+FLUD	0.62					
0.54 (0.22 – 1.29)	0.67 (0.16 – 2.62)	0.89 (0.25 – 3.09)	FLUD+TBI	0.59				
0.51 (0.11 – 2.40)	0.65 (0.10 – 4.12)	0.86 (0.23 – 3.21)	0.97 (0.28 – 3.34)	ETP+TBI	0.46			
0.41 (0.12 – 1.38)	0.51 (0.10 – 2.50)	0.67 (0.26 – 1.68)	0.75 (0.33 – 1.77)	0.79 (0.30 – 2.00)	BU+CY	0.31		
0.33 (0.10 – 1.08)	0.42 (0.08 – 1.96)	0.56 (0.20 – 1.49)	0.63 (0.29 – 1.34)	0.65 (0.23 – 1.77)	0.83 (0.57 – 1.19)	CY+TBI	0.18	
0.23 (0.07 – 0.71)	0.29 (0.06 – 1.35)	0.38 (0.09 – 1.58)	**0.43 (0.21 – 0.87)**	0.45 (0.11 – 1.88)	0.57 (0.19 – 1.67)	0.69 (0.25 – 1.91)	ATG+BU+FLUD	0.05
0.13 (0.03 – 0.62)	0.16 (0.02 – 1.04)	**0.22 (0.05 – 0.90)**	**0.24 (0.06 – 0.88)**	0.25 (0.06 – 1.09)	**0.32 (0.10 – 0.97)**	0.39 (0.13 – 1.11)	0.57 (0.13 – 2.44)	MELPH+TBI

Interventions are sorted from left to right in order of decreasing SUCRA value (i.e., preferred treatments appear first). For each comparison shown, the upper/left-most regimen is the comparator group while the lower/right-most treatment is the reference treatment; a value <1 suggests fewer cases of acute GVHD with the comparator than with the reference group. Statistically significant differences (i.e., estimates with a 95% credible interval excluding 1) are bolded and underlined. Comparisons with direct evidence, simple indirect evidence, and compound indirect evidence have been highlighted in red, orange, and yellow, respectively.

3.8. Findings: Chronic GVHD

All studies evaluating chronic GVHD outcomes reported the number of patients alive at 100 days (evaluable patients), or it was possible to approximate this number from survival curves. Analyses based on the number of evaluable patients were preferred by clinical experts and extensive cGVHD was considered more clinically relevant than overall cGVHD. Thus, findings from an NMA of extensive cGVHD in evaluable patients are presented below. The NMA for overall cGVHD in evaluable patients is reported in the supplement to the review.

Seven studies[27, 29, 31, 35, 45-47] evaluating 753 patients alive at 100 days had data available for the extensive cGVHD outcome. Eight conditioning regimens were evaluated in these studies; however, two of these interventions formed a disjoint pair and the remaining 6 interventions were distributed between 2 networks of 3 interventions each (**Figure 8**). Consultation with clinical experts determined the network of primary clinical relevance to be the one including CY+TBI, BU+CY, and BU+FLUD. This primary network included data from four studies[29, 35, 45, 47] and is the focus of the analyses below. The remaining 3-intervention network and disjoint pair of interventions have been summarized in the supplement to the review.

Figure 8: Network diagrams for the NMA of extensive cGVHD in evaluable patients

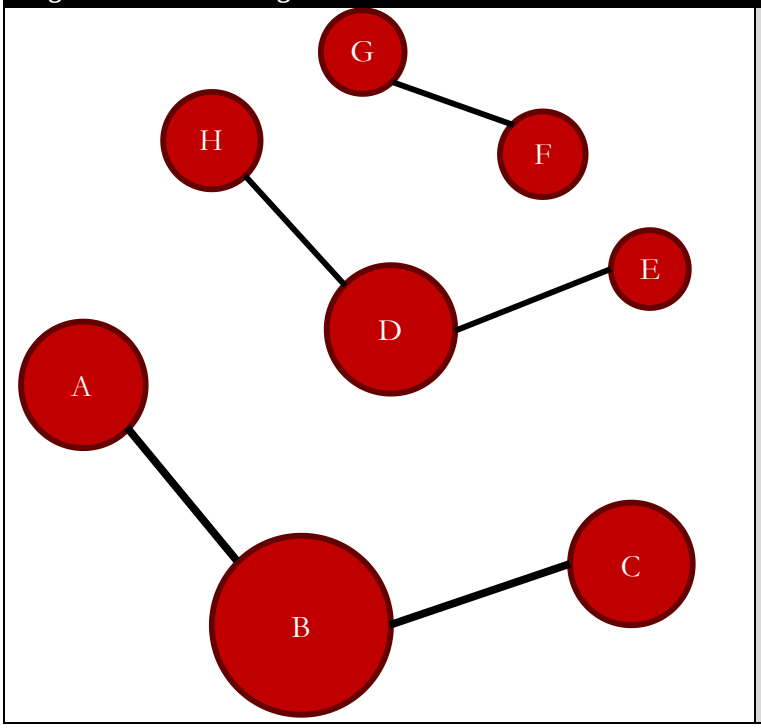


Figure legend
A = CY+TBI
B = BU+CY
C = BU+FLUD
D = FLUD+TBI
E = ATG+TLI
F = mBU+FLUD
G = mBU+CY
H = ATG+BU+FLUD

ATG = anti-thymocyte globulin
BU = busulfan
CY = cyclophosphamide
FLUD = fludarabine
mBU+CY = BU+CY+ hydroxyurea+semustine+cytosine arabinoside
mBU+FLUD = BU+FLUD+hydroxyurea+semustine+cytosine arabinoside
TBI = total body irradiation
TLI = total lymphoid irradiation

3.8.1. Results from traditional pairwise meta-analyses

For the primary network evaluating extensive cGVHD in patients alive at 100 days, **Table 11** presents summaries of pairwise estimates that were derived from direct evidence (i.e., the head-to-head trials), alongside the numbers of available trials and patients for each direct comparison, as well as the related summary estimate derived from the network meta-analysis.

The two comparisons with direct evidence were each informed by 2 studies. Pairwise meta-analysis suggested there were no significant differences in conditioning regimens with respect to reduction of risk of extensive cGVHD in patients alive at 100 days post-transplant (**Table 11**). Estimates from traditional pairwise meta-analyses compared well with estimates derived from network meta-analysis.

Table 5: Summary of results from meta-analysis and NMA, Extensive cGVHD in evaluable patients					
Comparison		# of Trials (patients)	Heterogeneity (I^2)	Direct Estimate OR (95% CrI)	FE NMA Estimate OR (95% CI)
				*Values <1 favor comparator	
Comparator	Reference				
BU+FLUD	BU+CY	2 (209)	0.000	0.74 (0.39–1.38)	0.73 (0.39–1.39)
CY+TBI	BU+CY	2 (225)	0.000	0.89 (0.51–1.53)	0.88 (0.51–1.52)

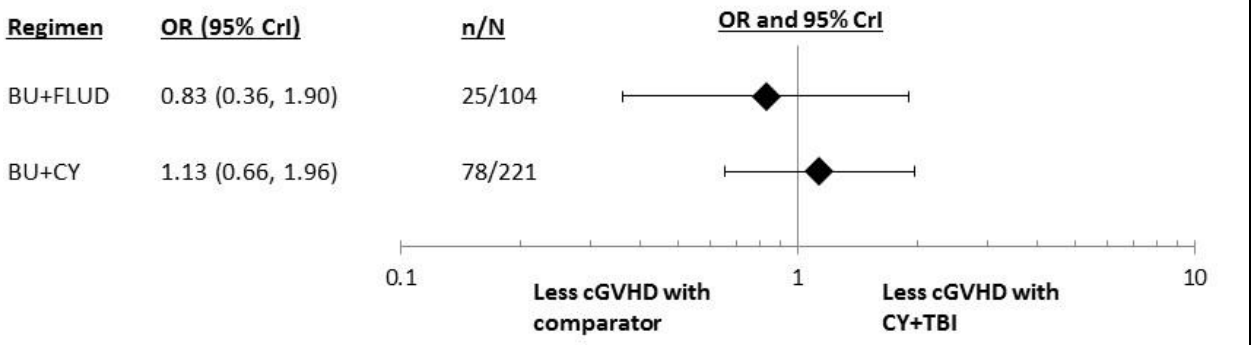
3.8.2. Results from network meta-analysis

Two of the three comparisons in the network were informed by direct evidence, both of which were informed by 2 studies each and approximately 200 patients. Model fit statistics for the NMA indicated that both the fixed-effects (FE) and random-effects (RE) models had an adequate fit. Posterior residual deviance values for the FE and RE models of 6.985 and 7.45, respectively, were obtained, both of which demonstrated a desirable fit, given the 8 data points in the model. DIC values (46.822 and 48.426, respectively) suggested similar fit. For reasons described earlier, the FE model was preferred and is the focus in describing results in the main text, while full results from the RE analysis are provided in the review supplement.

3.8.2.1. Comparisons versus CY+TBI

Figure 9 presents a forest plot summarizing comparisons of all conditioning regimens in the evidence network to the chosen reference therapy, CY+TBI. All regimens were associated with credible intervals that included 1, and thus no important differences in risk of aGVHD compared to CY+TBI were identified. Interpretations from the RE analysis were analogous.

Figure 9: Forest plot of comparisons from NMA versus CY+TBI—Extensive cGVHD in evaluable patients

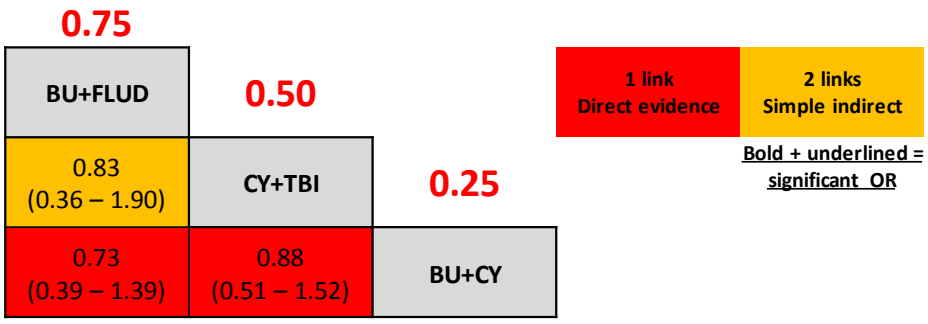


Odds ratios from network meta-analysis comparing each regimen to the reference regimen of CY+TBI are presented along with the number of events and patients per regimen. Odds ratios <1 suggest more risk of extensive chronic GVHD with CY+TBI while odds ratios >1 suggest less risk of extensive chronic GVHD with CY+TBI.

3.8.2.2. Comparisons between all conditioning regimens

Figure 10 presents a league table of the estimates for all 3 pairwise comparisons generated from FE NMA. There were no statistically significant differences between conditioning regimens with respect to the risk of extensive cGVHD in patients that survived to 100 days; findings from the RE analysis resulted in analogous clinical interpretations.

Figure 10: League table summary of pairwise comparisons from FE NMA, Odds Ratios with 95% CrI, Extensive cGVHD in evaluable patients



Interventions are sorted from left to right in order of decreasing SUCRA value (i.e., preferred treatments appear first). For each comparison shown, the upper/left-most regimen is the comparator group while the lower/right-most treatment is the reference treatment; a value <1 suggests fewer cases of extensive chronic GVHD with the comparator than with the reference group. Statistically significant differences (i.e., estimates with a 95% credible interval excluding 1) are bolded and underlined. Comparisons with direct evidence and simple indirect evidence have been highlighted in red and orange, respectively.

3.9. Findings: Disease Relapse at 2–3 years

Fourteen studies reported on the relapse of patients, with follow-up times ranging between 0.3–9 years. Ten studies[26, 29, 32, 35, 37, 44-46, 48, 51] evaluating 1,317 patients reported a mean or median follow-up time between 2–3 years, and based upon data availability this follow-up period was selected for analysis. Ten conditioning regimens were evaluated in these studies (**Figure 11**), however 6 of these interventions formed disjoint pairs, leaving the remaining 4 interventions for NMA. This 4-intervention network including data from seven studies[29, 32, 35, 37, 45, 48, 51] was selected as the focus of the analyses detailed below. Clinical findings related to the 3 disjoint pairs of conditioning regimens have been summarized in the supplement to the report.

Figure 11: Network diagrams for the NMA of relapse after 2–3 years

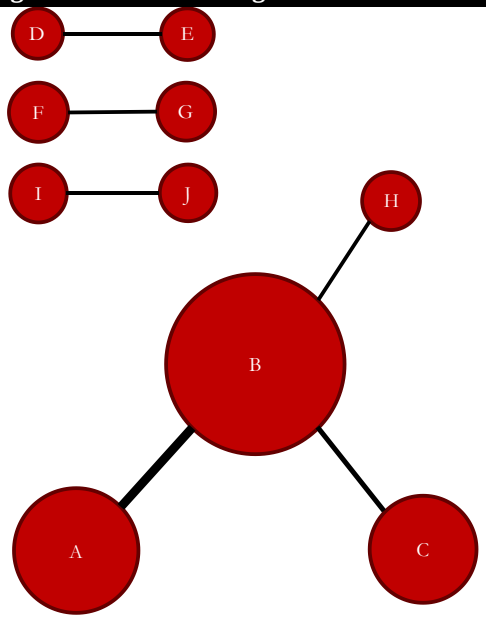


Figure legend
A = CY+TBI
B = BU+CY
C = BU+FLUD
D = FLUD+TBI
E = TBI
F = ALZ+MELPH+FLUD
G = ATG+THIO+MELPH+CY
H = ETP+TBI
I = mBU+CY
J = mBU+FLUD

ATG = anti-thymocyte globulin; ALZ = alemtuzumab; BU = busulfan; CY = cyclophosphamide; ETP = etoposide; FLUD = fludarabine; mBU+CY = BU+CY+ hydroxyurea +semustine+cytosine arabinoside; mBU+FLUD = BU+FLUD+hydroxyurea+semustine+cytosine arabinoside; MELPH = melphalan; TBI = total body irradiation; THIO = thiotepa

3.9.1. Results from traditional pairwise meta-analyses

For the network evaluating relapse at 2–3 years of follow-up, **Table 11** presents summaries of pairwise estimates that were derived from direct evidence (i.e., the head-to-head trials), alongside the numbers of available trials and patients for each direct comparison, as well as the related summary estimate derived from the network meta-analysis.

The direct comparison with the largest number of studies available was BU+CY vs. CY+TBI (4 studies, 530 patients), followed by BU+CY vs. BU+FLUD (2 studies, 371 patients). There were no significant differences identified by the pairwise comparisons of conditioning regimens with respect to effects on 2-3 year relapse. Estimates from traditional pairwise meta-analyses compared well with estimates derived from network meta-analysis.

Table 6: Summary of results from meta-analysis and NMA, Relapse, 2–3 years

Comparison		# of Trials (patients)	Heterogeneity (I^2)	Direct Estimate OR (95% CrI)	FE NMA Estimate OR (95% CI)
				**Values <1 favor comparator*	
Comparator	**Reference**				
BU+CY	CY+TBI	4 (530)	66.476	0.77 (0.48–1.23)	0.74 (0.47–1.17)
BU+CY	ETP+TBI	1 (114)	—	0.71 (0.34–1.52)	0.71 (0.33–1.51)
BU+CY	BU+FLUD	2 (371)	0.000	0.71 (0.45–1.12)	0.71 (0.45–1.12)

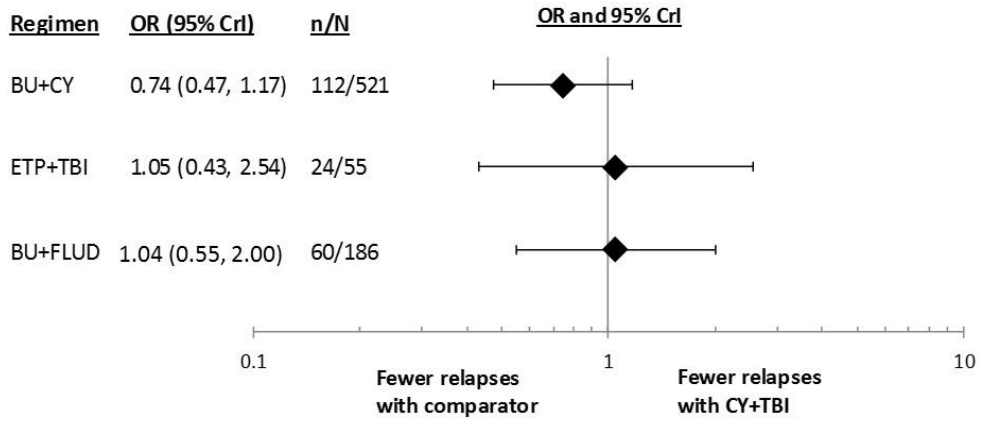
3.9.2. Results from network meta-analysis

Three of the six possible comparisons in the network were informed by direct evidence, with two of these comparisons being informed by more than one study. Model fit statistics for the NMA indicated that both the fixed-effects (FE) and random-effects (RE) models had an adequate fit. Posterior residual deviance values for the FE and RE models of 21.31 and 15.22, respectively, were obtained, both of which demonstrated a desirable fit, given the 14 data points in the model. DIC values (90.729 and 88.062, respectively) suggested similar fit; a FE model was preferred for reasons described earlier.

3.9.2.1. Comparisons versus CY+TBI

Figure 12 presents a forest plot summarizing comparisons of all conditioning regimens in the evidence network to the chosen reference therapy, CY+TBI. All regimens were associated with credible intervals that included 1 and, thus, were not significantly different from CY+TBI in preventing relapse within 2–3 years of transplant.

Figure 12: Forest plot of comparisons from NMA versus CY+TBI—Relapse, 2–3 years

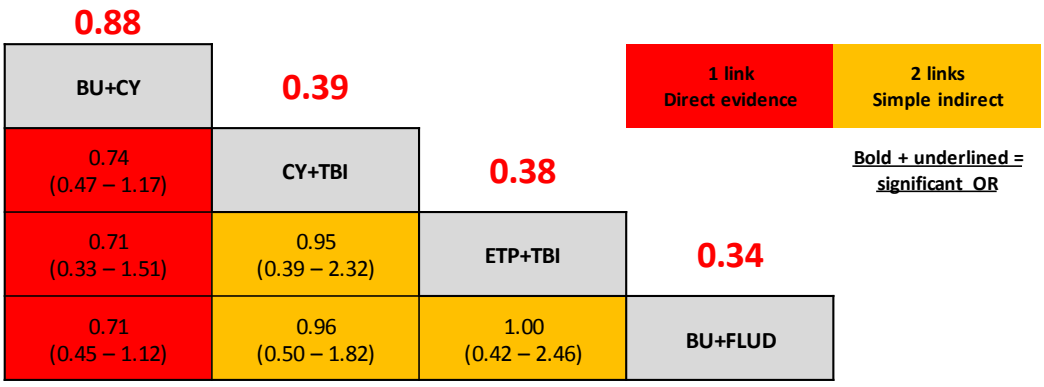


Odds ratios from network meta-analysis comparing each regimen to the reference regimen of CY+TBI are presented along with the number of events and patients per regimen. Odds ratios <1 suggest more risk of relapse with CY+TBI while odds ratios >1 suggest less risk of relapse with CY+TBI.

3.9.2.2. Comparisons between all conditioning regimens

Figure 13 presents a league table of the estimates for all pairwise comparisons generated by NMA. There were no significant differences between conditioning regimens with respect to the risk of relapse within 2–3 years of transplant.

Figure 13: League table summary of pairwise comparisons from NMA, Odds Ratios with 95% CrI, Relapse, 2–3 years



Interventions are sorted from left to right in order of decreasing SUCRA value (i.e., preferred treatments appear first). For each comparison shown, the upper/left-most regimen is the comparator group while the lower/right-most treatment is the reference treatment; a value <1 suggests fewer relapses with the comparator than with the reference group. Statistically significant differences (i.e., estimates with a 95% credible interval excluding 1) are bolded and underlined. Comparisons with direct evidence and simple indirect evidence have been highlighted in red and orange, respectively.

3.10. Findings: Relapse, follow-up <2 years or ≥4 years

Five studies[27, 31, 35, 41, 47] reported relapse data that were not in the follow-up period of 2–3 years and, therefore, could not be included in the NMA reported above. Seven conditioning regimens were evaluated, with follow-up times of 0.3 years[35] and 4–9 years[27, 31, 35, 41, 47]. The data from these studies are summarized in **Table 12**. Calculation of pairwise comparison for each study identified two significant differences in the risk of relapse. At 4 years post-transplant, FLUD+TBI was associated with significantly fewer occurrences of relapse than ATG+TLI (OR = 0.28; CI = 0.11–0.67) and at 5 years post-transplant, BU+FLUD+ATG was associated with significantly fewer occurrences of relapse than FLUD+TBI (OR =

0.32; CI = 0.16–0.65). Given the similarity in follow-up times of the two studies, an NMA was conducted on these two studies alone using a FE model (233 patients included). The indirect estimate comparing BU+FLUD+ATG to the reference ATG+TLI demonstrated a significant difference between the two regimens (OR = 0.08; CrI = 0.03–0.26). The direct estimates for the other two comparisons in the NMA were similar to the ORs and CIs presented in Table 12. A league table of these findings is presented in **Figure 14**.

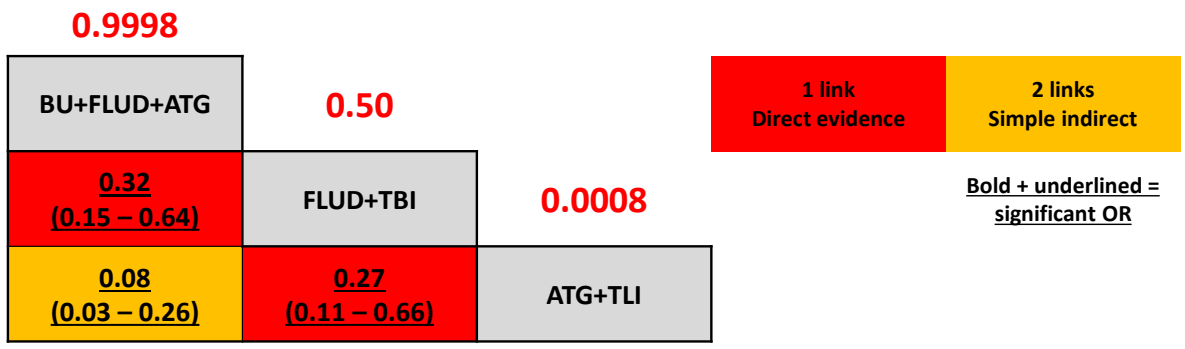
At 5 years post-transplant, no significant difference in relapse between BU+CY and BU+FLUD was identified by Liu et al.[47]. However, when only AML patients were analysed, the authors reported a significant difference in the two regimens (RR = 2.22 (95% CI: 1.15–4.29); p = 0.018; see Appendices).

Table 7: Summary of results: Risk of relapse <2 years or ≥4 years post-transplant

Author	Follow-up	Treatment regimen	Group risk	OR (95% CI)
Baron (2015)[27, 28]	4 years	FLUD+TBI	11/49 (22%)	**0.28 (0.11–0.67)**
		ATG+TLI	23/45 (50%)*	
Blaise (2013)[31]	5 years	BU+FLUD+ATG	19/69 (27%)	**0.32 (0.16–0.65)**
		FLUD+TBI	38/70 (54%)*	
Clift (1994)[35, 36]	0.3 years	CY+TBI	9/69 (13%)*	0.29 (0.07–1.10)
		BU+CY	3/73 (4%)	
Clift (1994)[35, 36]	9 years	CY+TBI	15/69 (22%)*	0.85 (0.38–1.93)
		BU+CY	14/73 (19%)	
Helenglass (1988)[41]	4.6 years	CY+TBI	7/36 (19%)*	0.16 (0.02–1.38)
		MELPH+TBI	1/27 (3.7%)	
Liu (2013)[47]	5 years	BU+CY	7/54 (13%)	0.86 (0.29–2.55)
		BU+FLUD	8/54 (15%)*	

*denotes the reference group for each pairwise comparison
ATG = anti-thymocyte globulin; BU = busulfan; CY = cyclophosphamide; FLUD = fludarabine; MELPH = melphalan; TBI = total body irradiation; TLI = total lymphoid irradiation

Figure 14: League table summary of pairwise comparisons from NMA, Odds Ratios with 95% CrI, Relapse, 4–5 years



Interventions are sorted from left to right in order of decreasing SUCRA value (i.e., preferred treatments appear first). For each comparison shown, the upper/left-most regimen is the comparator group while the lower/right-most treatment is the reference treatment; a value <1 suggests fewer relapses with the comparator than with the reference group. Statistically significant differences (i.e., estimates with a 95% credible interval excluding 1) are bolded and underlined. Comparisons with direct evidence and simple indirect evidence have been highlighted in red and orange, respectively.

3.11. Findings: Veno-occlusive disease or hepatic dysfunction, ≥100-day follow-up

Five studies[29, 37, 45, 46, 51] reported data on the occurrence of VOD and two studies[35, 41] reported hepatic dysfunction, with follow-up times ranging from 28 days to over 9 years. Hepatic dysfunction was defined as bilirubin elevated to >2 mg/dl (i.e., >30 μmol/l). Consultation with clinical experts suggested that VOD would be expected to occur before 100 days post-transplant; thus, only studies evaluating patients for at least 100 days were included in the NMA (5 studies; 562 patients) and the two remaining studies that reported VOD at 28 days post-transplant have been summarized narratively. Six conditioning regimens were evaluated in the 5 studies available for NMA **(Figure 15**); however, 2 of these interventions formed disjoint pairs, leaving the remaining 4 interventions for NMA. This 4-intervention network containing data from 4 studies[29, 41, 45, 51] is the focus of the analyses below. The disjoint pair of conditioning regimens has been summarized in the appendices to the report.

Figure 15: Network diagrams for the NMA of VOD or hepatic dysfunction

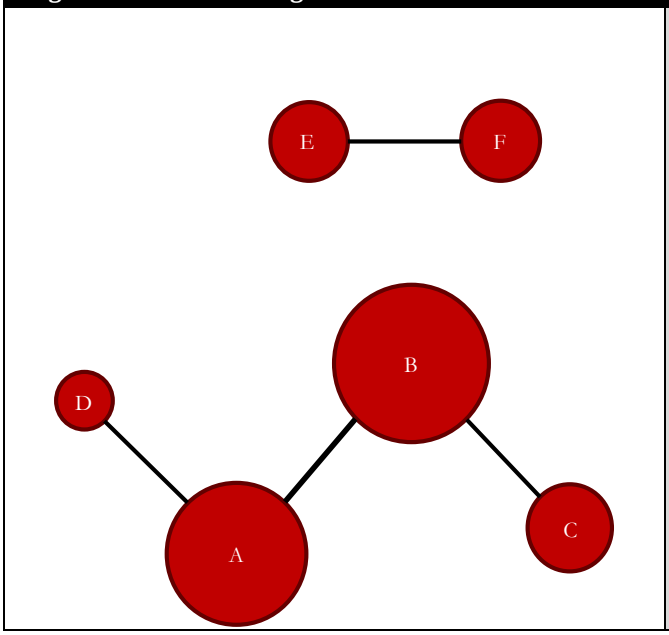


Figure legend
A = CY+TBI
B = BU+CY
C = BU+FLUD
D = MELPH+TBI
E = mBU+FLUD
F = mBU+CY

BU = busulfan
CY = cyclophosphamide
FLUD = fludarabine
mBU+CY = BU+CY+ hydroxyurea+semustine+cytosine arabinoside
mBU+FLUD = BU+FLUD+hydroxyurea+semustine+cytosine arabinoside
MELPH = melphalan
TBI = total body irradiation

3.11.1. Results from traditional pairwise meta-analyses

For the 4-intervention network, **Table 13** presents summaries of pairwise estimates that were derived from direct evidence (i.e., the head-to-head trials), alongside the numbers of available trials and patients for each direct comparison, as well as the related summary estimate derived from the network meta-analysis.

The direct comparison with the largest number of studies available was CY+TBI vs. BU+CY (2 studies, 268 patients). Significantly fewer patients developed VOD or hepatic dysfunction when CY+TBI was used as a conditioning regimen than BU+CY. There were no significant differences between BU+FLUD and BU+CY. Estimates from traditional pairwise meta-analyses compared well with estimates derived from network meta-analysis.

Table 8: Summary of results from meta-analysis and NMA, VOD or hepatic dysfunction, ≥100 days follow-up

Comparison		# of Trials (patients)	Heterogeneity (I^2)	Direct Estimate OR (95% CrI)	FE NMA Estimate OR (95% CrI)
				**Values <1 favor comparator*	
Comparator	**Reference**				
CY+TBI	MELPH+TBI	1 (63)	—	0.54 (0.19–1.50)	0.53 (0.19–1.47)
CY+TBI	BU+CY	2 (268)	0.000	**0.19 (0.05–0.70)**	**0.15 (0.03–0.48)**
BU+FLUD	BU+CY	1 (126)	—	0.41 (0.10–1.68)	0.39 (0.07–1.49)

3.11.2. Results from network meta-analysis

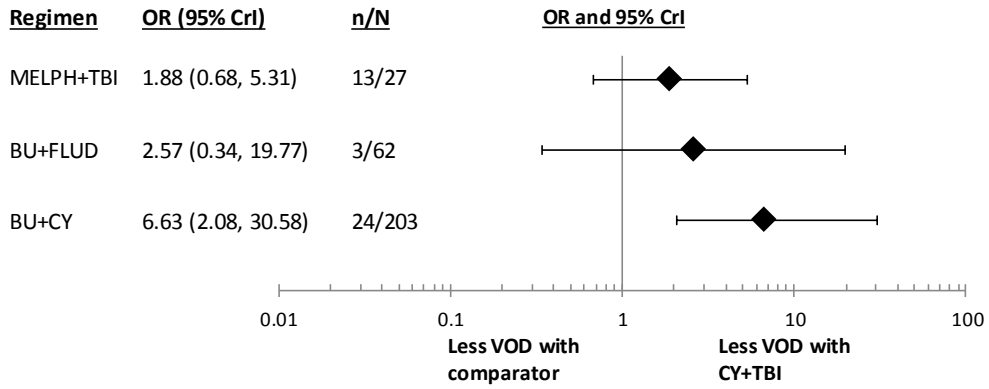
Three of the six possible comparisons in the network were informed by direct evidence, with one of these comparisons being informed by two studies. Model fit statistics for the NMA indicated that both the fixed-effects (FE) and random-effects (RE) models had an adequate fit. Posterior residual deviance values

for the FE and RE models of 8.11 and 7.97, respectively, were obtained, both of which demonstrated a desirable fit, given the 8 data points in the model. DIC values (41.595 and 41.881, respectively) suggested similar fit; a FE model was preferred due to the number of single-study connections.

3.11.2.1. Comparisons versus CY+TBI

Figure 16 presents a forest plot summarizing comparisons of all conditioning regimens in the evidence network to the chosen reference therapy, CY+TBI. CY+TBI was significantly better than BU+CY in preventing VOD or hepatic dysfunction after transplant.

Figure 16: Forest plot of comparisons from NMA versus CY+TBI—VOD or hepatic dysfunction, ≥100 days follow-up

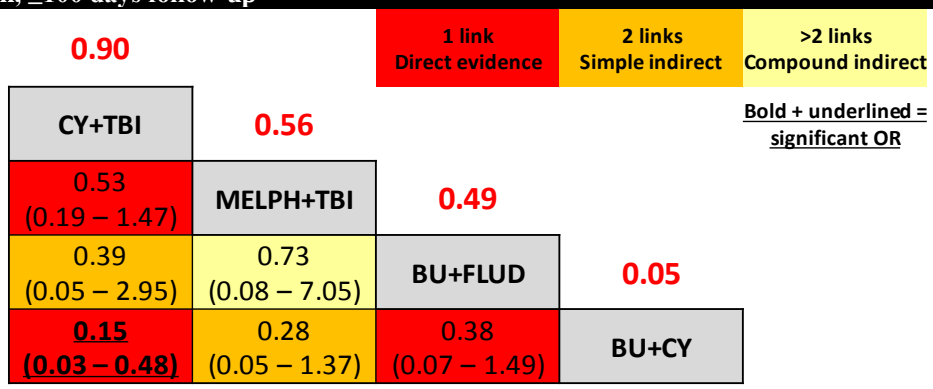


Odds ratios from network meta-analysis comparing each regimen to the reference regimen of CY+TBI are presented along with the number of events and patients per regimen. Odds ratios <1 suggest more risk of VOD or hepatic dysfunction with CY+TBI while odds ratios >1 suggest less risk of VOD or hepatic dysfunction with CY+TBI.

3.11.2.2. Comparisons between all conditioning regimens

Figure 17 presents a league table of the estimates for all pairwise comparisons generated by NMA. CY+TBI was associated with significantly fewer cases of VOD and hepatic dysfunction when compared to BU+CY. Otherwise, no comparisons of conditioning regimens demonstrated significant differences.

Figure 17: League table summary of pairwise comparisons from FE NMA, Odds Ratios with 95% CrI, VOD or hepatic dysfunction, ≥100 days follow-up



Interventions are sorted from left to right in order of decreasing SUCRA value (i.e., preferred treatments appear first). For each comparison shown, the upper/left-most regimen is the comparator group while the lower/right-most treatment is the reference treatment; a value <1 suggests fewer cases of VOD or hepatic dysfunction with the comparator than with the reference group. Statistically significant differences (i.e., estimates with a 95% credible interval excluding 1) are bolded and underlined. Comparisons with direct evidence and simple indirect evidence have been highlighted in red and orange, respectively.

3.12. Findings: Veno-occlusive disease or hepatic dysfunction, 28-day follow-up

Two studies[35, 37] reported relapse data with a follow-up period under 100 days and, therefore, could not be included in the NMA reported above. Both studies evaluated CY+TBI compared to BU+CY at 28 days post-transplant. A pairwise meta-analysis was conducted on these two studies (262 patients). At 28 days post-transplant, there was no significant difference in the risk of VOD or hepatic dysfunction between CY+TBI and BU+CY (OR = 1.02; 95% CI = 0.54–1.93). This is in contrast to the findings for ≥100 days of follow-up in which the risk of VOD was reduced significantly by CY+TBI when compared to BU+CY (OR = 0.15; CrI = 0.03–0.48).

Table 9: Summary of results: VOD or hepatic dysfunction, 28-day follow-up

Author	Outcome definition	Follow-up	Treatment regimen	Group risk	Study-level OR (95% CI)	Heterogeneity (I^2)	Meta-analysis OR (95% CI)
Clift (1994)[35, 36]	Hepatic dysfunction	28 days	CY+TBI	48/69 (70%)*	1.01 (0.50–2.08)	0.000	1.02 (0.54–1.93)
			BU+CY	51/72 (71%)			
Devergie (1995)[37]	VOD	28 days	CY+TBI	4/55 (7%)*	1.06 (0.27–4.17)		
			BU+CY	5/65 (8%)			

*denotes the reference group for each pairwise comparison
BU = busulfan; CY = cyclophosphamide; TBI = total body irradiation; VOD = veno-occlusive disease

3.13. Findings: Bronchiolitis obliterans

One study reported data on the risk of bronchiolitis obliterans[49-51]. In 112 patients that were followed from 1.7 to 9.3 years after transplantation, those randomized to CY+TBI (3/58 patients; 5%) were significantly less likely to experience bronchiolitis obliterans than those randomized to BU+CY (14/54 patients; 26%) (OR = 0.16; 95% CI = 0.04–0.58).

3.14. Findings: Specific harms

Few studies reported harm-related data. As a result, these data have been summarized narratively under the headings for each specific harm of interest. Where possible, ORs have been calculated to identify significant differences between pairs of interventions.

3.14.1. Total or serious adverse events

No studies reported total or serious adverse events.

3.14.2. Infection—Sepsis/bacterial infection

Two studies[29, 35] reported data on sepsis specifically, and two other studies[27, 44] reported data on bacterial infection. The findings from these studies are presented in **Table 15**.

Clift et al.[35] demonstrated that CY+TBI was associated with a significantly increased odds of sepsis at 100 days; however, Blaise et al.[29] did not find this difference to be significant at a mean of 1.9 years of follow-up. No other significant findings were found.

Table 10: Summary of results: Risk of sepsis or bacterial infection in the evaluation of conditioning regimens

Author	Follow-up	Outcome definition	Conditioning regimen	Group risk	OR
Blaise (1992)[29, 30]	Mean 1.9 +/- 0.9 years	Bacteremia	CY+TBI	12/50 (24%)	1.02 (0.41–2.57)
			BU+CY	12/51 (24%)*	
Clift (1994)[35, 36]	0.3 years (100 days)	At least one positive blood culture	CY+TBI	25/69 (36%)	**2.62 (1.20–5.69)**
			BU+CY	13/73 (18%)*	
Kornblit (2013)[44]	3 years	Bacterial infection	FLUD+TBI	28/41 (68%)	1.23 (0.50–3.03)
			TBI	28/44 (64%)*	
Baron (2015)[27, 28]	0.3 years (100 days)	Bacterial infection	FLUD+TBI	19/49 (39%)*	1.97 (0.87–4.49)
			ATG+TLI	25/45 (56%)	

*denotes the reference group for each pairwise comparison
ATG = anti-thymocyte globulin; BU = busulfan; CY = cyclophosphamide; FLUD = fludarabine; TBI = total body irradiation; TLI = total lymphoid irradiation

3.14.3. Infection—Pneumonia/interstitial pneumonitis

Six studies[29, 37, 43, 45, 46, 51] reported various types of pneumonia occurring during follow-up, including severe pneumonia (as per the diagnostic criteria of the American Thoracic Society); bacterial, fungal, or *Pneumocystis carinii* pneumonia; interstitial pneumonitis; and pneumonia or interstitial pneumonitis. The findings from these studies are presented in **Table 16**.

The risk of severe pneumonia at 1.4 and 3.3 years of longest follow-up was significantly higher in the modified BU+FLUD group compared to the modified BU+CY group[46]. This study was suspended 1.4 years after initiation based on the significant difference in the risk of severe pneumonia between the two conditioning regimens. The difference continued to be significant after 3.3 years; however, the risk of all

types and severities of pneumonia was not significantly different between the two treatment groups. There were no other significant differences in any of the compared conditioning regimens with respect to the risk of pneumonia or interstitial pneumonitis.

Table 11: Summary of results: Risk of pneumonia in the evaluation of conditioning regimens

Author	Follow-up	Outcome definition	Treatment regimen	Group risk	OR
Liu (2013)[46]	≤1.4 years	Severe pneumonia[a]	mBU+FLUD	10/52 (19%)	**3.97 (1.02–15.37)**
			mBU+CY	3/53 (6%)*	
Liu (2013)[46]	≤3.3 years	Severe pneumonia[a]	mBU+FLUD	14/52 (27%)	**3.54 (1.17–10.70)**
			mBU+CY	5/53 (9%)*	
Liu (2013)[46]	≤3.3 years	Bacterial, fungal, or *Pneumocystis carinii* pneumonia	mBU+FLUD	8/52 (15%)	0.89 (0.31–2.51)
			mBU+CY	9/53 (17%)*	
Blaise (1992)[29, 30]	Mean 1.9 +/- 0.9 years (0.5 years minimum)	Interstitial pneumonitis	CY+TBI	5/50 (10%)	2.72 (0.50–14.74)
			BU+CY	2/51 (4%)*	
Devergie (1995)[37]	28 days	Interstitial pneumonitis	CY+TBI	12/55 (22%)	1.37 (0.55–3.41)
			BU+CY	11/65 (17%)*	
Kim (2012)[42, 43]	Median 1.5 (0.1–7.1) years	Pneumonia/ pneumonitis	CY+ATG	6/40 (15%)	3.62 (0.69–19.10)
			CY+ATG+FLUD	2/43 (5%)*	
Lee (2013)[45]	Median 2.2 (0.3–4.7) years	Interstitial pneumonitis	BU+CY	3/64 (5%)	1.48 (0.24–9.15)
			BU+FLUD	2/62 (3%)*	
Ringdén (1999)[49-51]	1.7–9.25 years	Interstitial pneumonitis	CY+TBI	13/79 (16%)*	1.13 (0.50–2.52)
			BU+CY	16/88 (18%)	

[a]According to diagnostic criteria of the American Thoracic Society, 2001
*denotes the reference group for each pairwise comparison
ATG = anti-thymocyte globulin; BU = busulfan; CY = cyclophosphamide; FLUD = fludarabine; mBU+CY = BU+CY+ hydroxyurea+semustine+cytosine arabinoside; mBU+FLUD = BU+FLUD+hydroxyurea+semustine+cytosine arabinoside; TBI = total body irradiation

3.14.4. Infection—CMV

Six studies[27, 35, 43, 45, 46, 51] reported on the risk of CMV antigenemia or reactivation, CMV interstitial pneumonitis, or isolation of CMV from blood, urine, or throat swabs. The findings from these studies are presented in **Table 17**. None of the studies identified significant differences between the conditioning regimens with respect to CMV infection outcomes.

Table 12: Summary of results: Risk of CMV in the evaluation of conditioning regimens

Author	Follow-up	Outcome definition	Treatment regimen	Group risk	OR (95% CI)
Liu (2013)[46]	3.3 years	CMV antigenemia	mBU+FLUD	3/52 (6%)*	2.49 (0.61–10.19)
			mBU+CY	7/53 (13%)	
Clift (1994)[35, 36]	100 days	CMV isolation from blood, urine, or throat swabs	CY+TBI	18/69 (23%)	1.26 (0.58–2.72)
			BU+CY	16/73 (22%)*	
Kim (2012)[42, 43]	Median 1.5 (0.1–7.1) years	CMV antigenemia	CY+ATG	24/40 (60%)	1.30 (0.55–3.12)
			CY+ATG+FLUD	23/43 (54%)*	

Table 12: Summary of results: Risk of CMV in the evaluation of conditioning regimens

Author	Follow-up	Outcome definition	Treatment regimen	Group risk	OR (95% CI)
Lee (2013)[45]	Median 2.2 (0.3–4.7) years	CMV antigenemia	BU+CY	40/64 (63%)	1.37 (0.67–2.80)
			BU+FLUD	34/62 (55%)*	
Ringdén (1999)[49-51]	1.7–9.25 years	CMV interstitial pneumonitis	CY+TBI	2/79 (3%)*	3.33 (0.67–16.51)
			BU+CY	7/88 (8%)	
Baron (2015)[27, 28]	100 days	CMV reactivation amongst CMV-seropositive patients	FLUD+TBI	11/36 (31%)*	2.13 (0.78–5.80)
			ATG+TLI	15/31 (47%)	

*denotes the reference group for each pairwise comparison. ATG = anti-thymocyte globulin; BU = busulfan; CY = cyclophosphamide; FLUD = fludarabine; mBU+CY = BU+CY+ hydroxyurea+semustine+cytosine arabinoside; mBU+FLUD = BU+FLUD+hydroxyurea+semustine+cytosine arabinoside; TBI = total body irradiation; TLI = total lymphoid irradiation

3.14.5. Infection—Fungal

Two studies[27, 44] reported data on fungal infections. At 100 days of follow-up, there was no significant difference in the risk of fungal infections in patients randomized to FLUD+TBI (3/45, 6%) or ATG+TLI (7/45, 16%) (OR = 2.58; 95% CI = 0.62–10.69)[27, 28]. Similarly, at 3 years of follow-up, there was no significant difference between FLUD+TBI (9/41, 22%) and TBI alone (9/44, 21%) in the risk of fungal infections (OR = 1.09; 95% CI = 0.39–3.10)[44].

3.14.6. Infection—Other

Other infectious outcomes were reported by two studies[44, 46] and included overall infections, viral infections, infection of skin and soft tissues, CNS infections, neutropenic fever, and post-transplantation lymphoproliferative disease. These data have been summarized in **Table 18**. There were no significant differences in any of the conditioning regimen comparisons for other infectious outcomes.

Table 13: Summary of results: Risk of other infectious outcomes in the evaluation of conditioning regimens

Author	Follow-up	Outcome definition	Treatment regimen	Group risk	OR
Liu (2013)[46]	3.3 years	Overall infections	mBU+FLUD	31/52 (60%)	1.32 (0.61–2.86)
			mBU+CY	28/53 (53%)*	
Liu (2013)[46]	3.3 years	CNS infection	mBU+FLUD	1/52 (2%)	Not estimable
			mBU+CY	0/53 (0%)*	
Liu (2013)[46]	3.3 years	Infection of skin or soft tissues	mBU+FLUD	4/52 (8%)	1.02 (0.24–4.32)
			mBU+CY	4/53 (8%)*	
Liu (2013)[46]	3.3 years	Neutropenic fever	mBU+FLUD	16/52 (31%)	1.13 (0.49–2.61)
			mBU+CY	15/53 (28%)*	
Liu (2013)[46]	3.3 years	Post-transplantation lymphoproliferative disease	mBU+FLUD	2/52 (4%)	Not estimable
			mBU+CY	0/53 (0%)*	
Kornblit (2013)[44]	3 years	Viral infection	FLUD+TBI	28/41 (68%)	1.49 (0.61–3.63)
			TBI	26/44 (60%)*	

*denotes the reference group for each pairwise comparison
BU = busulfan; CY = cyclophosphamide; FLUD = fludarabine; mBU+CY = BU+CY+ hydroxyurea+semustine+cytosine arabinoside; mBU+FLUD = BU+FLUD+hydroxyurea+semustine+cytosine arabinoside; TBI = total body irradiation

3.14.7. Other harms

Four studies[35, 41, 43, 47] reported harms other than those summarized above, including regimen-related toxicity, renal insufficiency, and pulmonary complications. These data have been summarized in **Table 19**.

The addition of fludarabine to the regimen CY+ATG was protective of regimen-related toxicity as well as pulmonary complications. The odds of regimen-related toxicity were 4.03 times higher (95% CI = 1.57–10.35) in the CY+ATG group than the CY+ATG+FLUD group. The odds of pulmonary complications were 4.09 times higher (95% CI = 1.13–12.75) in the CY+ATG group than the CY+ATG+FLUD group. Otherwise, there were no significant differences between the conditioning regimens tested. Seventeen percent of patients receiving BU+FLUD experienced Grade III–IV regimen-related toxicity compared to none of those receiving BU+CY. An odds ratio could not be estimated due to the zero events in the BU+CY group; however, the risk difference was significant (RD = 16.7%; 95% CI = 6.73–26.61). No other comparisons demonstrated significant differences between conditioning regimens.

Table 14: Summary of results: Risk of other harms in the evaluation of conditioning regimens

Author	Follow-up	Outcome definition	Treatment regimen	Group risk	OR
Clift (1994)[35, 36]	28 days	Renal insufficiency[a]	CY+TBI	4/69 (5.8%)	4.43 (0.48–40.67)
			BU+CY	1/73 (1.4%)*	
Helenglass (1988)[41]	5.4 years (minimum 1 year)	Renal insufficiency[b]	CY+TBI	9/36 (25%)*	1.5 (0.50–4.50)
			MELPH+TBI	9/27 (33%)	
Kim (2012)[42, 43]	Median 1.5 (0.1–7.1) years	Regimen-related toxicity[c]	CY+ATG	22/40 (55%)	**4.03 (1.57–10.35)**
			CY+ATG+FLUD	10/43 (23%)*	
Kim (2012)[42, 43]	Median 1.5 (0.1–7.1) years	Pulmonary complications[d]	CY+ATG	14/40 (35%)	**4.09 (1.13–12.75)**
			CY+ATG+FLUD	5/43 (12%)*	
Liu (2013)[47]	Median 1.7 (0.01–5.8) years	Regimen-related toxicity[e]	BU+CY	51/54 (94%)	3.86 (1.00–14.93)
			BU+FLUD	44/54 (82%)*	
Liu (2013)[47]	Median 1.7 (0.01–5.8) years	Regimen-related toxicity[f]	BU+CY	9/54 (17%)	Not estimable
			BU+FLUD	0/54 (0%)	

[a]Creatinine >3.0 mg/dl
[b]Creatinine >300 µmol/l on 3 consecutive occasions
[c]Any regimen-related toxicity, including VOD, hematuria, pulmonary complications, and death from any cause
[d]Including bacterial pneumonia, pulmonary TB reactivation, pneumonitis caused by aGVHD, pulmonary edema caused by acute renal failure, and other unknown causes of pulmonary complications
[e]Any regimen-related toxicity
[f]Grades III-IV regimen-related toxicity
*denotes the reference group for each pairwise comparison
ATG = anti-thymocyte globulin; BU = busulfan; CY = cyclophosphamide; FLUD = fludarabine; MELPH = melphalan; TBI = total body irradiation; VOD = veno-occlusive disease

3.15. Assessment of between-study heterogeneity and inconsistency

As indicated earlier in the report, several variations in patient and study characteristics were noted across the included studies in terms of types of patient indication, year of study publication and other aspects. Unfortunately, subgroup analyses and meta-regression analyses were not feasible in the context of this review due to the presence of many single-study connections across networks (which precludes performance of meaningful meta-regression analyses), as well as a failure of studies to report outcomes in subgroups of relevance which would enable the performance of more focused analyses. Given the high degree of heterogeneity present and the limited ability to evaluate its corresponding impact on findings, the validity and applicability of findings to specific populations must be considered carefully.

None of the networks of evidence for the outcomes studied in the current review included a closed loop; given this finding, formal analyses to evaluate the consistency assumption for network meta-analysis were not performed.

3.16. Discussion of findings

3.17. Summary of main findings

A variety of strategies for conditioning regimens in the population of patients undergoing aHSCT were identified in the context of the current review. Overall, a total of 15 trials that enrolled 1,913 patients were available for inclusion in network meta-analyses. Substantial variability in patient populations was noted with respect to age, underlying hematologic disease, disease risk of relapse/mortality, and transplant donor status (i.e., related vs. unrelated, matched vs. unmatched). Most comparisons between interventions in the analyses were based on indirect evidence only, and in several cases only single small studies were available. Further discussion of findings from the completed work follows below. *It is emphasized for readers that findings from network meta-analysis described in the report should be interpreted with caution given the high degree of heterogeneity present between studies which could not be well addressed based upon the evidence structure as well as the mixed reporting and eligibility criteria of eligible studies.* The numbers of patients studied were also limited, reflecting the challenges of conducting studies in this population.

NMAs of available mortality data were conducted at 100 day follow-up, as well as after 1, 2, 4 and 5 years. Amongst regimens that were compared, no important differences were identified within the first 100 days. For longer follow-up durations, there was evidence that CY+TBI may be associated with improved survival compared with both BU+CY and with BU+FLUD. NMAs for non-relapse mortality were not carried out due to reporting limitations regarding competing risks in several studies, and thus a qualitative inspection of study findings was performed; review of available study data did not identify definitive differences between regimens. A network meta-analysis involving four treatment regimens was also carried out to assess their relative effects in terms of incidence of disease relapse, and no significant differences between interventions were found. Network meta-analyses comparing totals of nine and three different regimens also failed to identify strong evidence of important differences between interventions in terms of the incidence of acute and chronic GVHD. A network meta-analysis for the occurrence of veno-occlusive disease within 100 days after transplant identified one significant difference, a reduced risk with CY+TBI relative to BU+CY. Overall, while the compilation of evidence networks provides insight for physicians and researchers in this realm, the ability to compare interventions was limited by the both the lack of studies as well as their diversity of comparisons and study populations. Many knowledge gaps remain and further studies are needed to further establish the relative benefits and harms of different conditioning regimens.

3.18. Strengths and limitations

Several limitations of the current review should be noted. First, there was considerable heterogeneity present across studies in terms of the populations recruited. This represents an ongoing difficulty of evidence syntheses in such transplant populations given the challenges of enrolling such patients; generalizability of findings can be difficult to assess, as the majority of studies are conducted in mixed populations of patients with a variety of malignancies, thereby complicating the application of summary findings to the task of treatment selection for individual patients. Variations in the standard care for conditioning regimens used to treatment patients around the world are broad, and thus in many cases we encountered niche regimens which may have been specific to individual institutions, thereby resulting in a high frequency of disconnected evidence networks. Additionally, supportive care administered to patients undergoing aHSCT continues to evolve rapidly over time, including changes in co-intervention administration (e.g. antibiotics, prophylactic GVHD regimens used) and changes in matching capability.

Several challenges from an evidence synthesis perspective should also be noted. Firstly, while study populations largely considered of a mixture of patients with assorted indications, outcome data for patient subgroups was often not reported, precluding the ability to compare findings within different populations. Furthermore, as discussed within the review, the relatively small number of related studies per treatment resulted in network structures which consisted of high numbers of comparisons informed by indirect evidence only (often of a compound nature) and networks formed primarily of single study connections between interventions; while fixed effects NMAs were thus a primary focus for our analyses, we have also presented all findings from random effects analyses in appendices to the report to respect the extent of heterogeneity present between studies and to provide readers with this information. Given these limitations in available evidence, the use of meta-regression to explore and account for the effects of heterogeneity was not feasible. The follow-up time at which endpoints were reported also varied amongst studies, thereby placing certain restrictions on the extent of the evidence that could by synthesized in some cases. The approach used to evaluate the occurrence of cGVHD also was variable across studies, sometimes precluding the ability to synthesize all study data.

3.19. Conclusions

Numerous drugs are used in a broad variety of single- and multi-agent conditioning regimens prior to HSCT. This review found that comparative evidence from randomized trials is lacking for many comparisons of these regimens, especially newer regimens, which has led to considerable practice variation between institutions. This systematic review of the evidence, incorporating network meta-analyses where possible, was conducted to address these gaps. While several gaps remain, certain points for clinical practice of interest to physicians were identified and have been noted in the interpretations above. Future studies of conditioning regimens should be stratified for competing factors, such as patient age and underlying disease, and donor factors. A standard arm from the existing evidence network should be included in future studies to leverage existing knowledge. Consistent outcome reporting is urgently needed in HSCT studies to improve network analysis and comparison with other studies.

Funding. This work was carried out with funds provided by a team grant from the Canadian Institutes of Health Research and the Drug Safety and Effectiveness Network.

References

1. Copeland E. Hematopoietic stem cell transplantation. New England Journal of Medicine. 2006;354:1813-26.

2. Bacigalupo A, Ballen K, Rizzo D, Giralt S, Lazarus H, Ho V, et al. Defining the Intensity of Conditioning Regimens:Working Definitions. Biol Blood Marrow Transplant. 2009;15:1628-33.

3. Gyurkocza B, Sandmeier B. Conditioning regimens for hematopoietic cell transplantation: one size does not fit all. Blood. 2014;124(3):344-53.

4. Iwamoto T, Hiraku Y, Oikawa S, Mizutani H, Kojima M, Kawanishi S. DNA intrastrand cross-link at the 5'-GA-3' sequence formed by busulfan and its role in the cytotoxic effect. Cancer Sci. 2004;95(5):454-8.

5. Hande K. Etoposide: four decades of development of a topoisomerase II inhibitor. Eur J Cancer. 1998;34(10):1514-21.

6. Ricci F, Tedeschi A, Morra E, Montillo M. Fludarabine in the treatment of chronic lymphocytic leukemia: a review. Ther Clin Risk Manag. 2009;5:187-207.

7. Samuels B, Bitran J. High-dose intravenous melphalan: a review. J Clin Oncol. 1995;13(7):1786-99.

8. Maanen M, Smeets C, Beijnen J. Chemistry, pharmacology and pharmacokinetics of N,N',N'' -triethylenethiophosphoramide (ThioTEPA). Cancer Treatment Rev. 2000;26(4):257-68.

9. Baskar R, Lee K, Yeo R, Yeoh KW. Cancer and Radiation Therapy: Current Advances and Future Directions. Int J Med Sci. 2012;9(3):193-9.

10. Hutton B, Salanti G, Caldwell D, Schmid C, Chaimani A, Cameron C, et al. The PRISMA Extension Statement for Reporting of Systematic Reviews Incorporating Network Meta-Analyses of Healthcare Interventions: Checklist and Explanations. Annals of Internal Medicine. 2015;In press.

11. Morrison A, Polisena J, Husereau D, Moulton K, et al. The effect of English-language restriction on systematic review-based meta-analyses: a systematic review of empirical studies. Int J Technol Assess Health Care. 2012;28(2):138-44.

12. Sampson M, McGowan J, Cogo E, Grimshaw J, Moher D, Lefebvre C. An evidence-based practice guideline for the peer review of electronic search strategies. Journal of Clinical Epidemiology. 2009;62(9):944-52.

13. Moher D, Liberati A, Tetzlaff J, Altman D, group P. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Medicine. 2008;6.

14. Higgins J, Altman D, Gotzsche P, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ. 2011;343.

15. Dias S, Sutton A, Welton N, Ades A. NICE DSU Technical Support Document 3: Heterogeneity: subgroups, meta-regression, bias and bias-adjustment 2012 [Available from: http://www.nicedsu.org.uk/TSD3%20Heterogeneity.final%20report.08.05.12.pdf].

16. Dias S, Welton N, Sutton A, Ades A. NICE DSU Technical Support Document 2:A Generalised Linear Modelling Framework for Pairwise and Network Meta-Analysis of Randomised Controlled Trials 2011 [Available from: http://www.nicedsu.org.uk/TSD2%20General%20meta%20analysis%20corrected%202Sep2016v2.pdf].

17. Dias S, Welton N, Sutton A, Caldwell D, Lu G, Ades A. NICE DSU Technical Support Document 4: Inconsistency in Networks of Evidence Based on Randomised Controlled Trials 2011 [Available from: http://www.nicedsu.org.uk/TSD4%20Inconsistency.final.15April2014.pdf].

18. Cameron C, Coyle D, Richter T, et al. Systematic review and network meta-analysis comparing antithrombotic agents for the prevention of stroke and major bleeding in patients with atrial fibrillation. BMJ Open. 2014;4(6).

19. Salanti G, Ades A, Ioannidis J. Graphical methods and numerical summaries for presenting results from multiple-treatment meta-analysis: an overview and tutorial. Journal of Clinical Epidemiology. 2011;64(2):163-71.

20. Salanti G. Indirect and mixed-treatment comparison, network, or multipletreatments meta-analysis: many names, many benefits, many concerns for the next generation evidence synthesis tool. Research Synthesis Methods. 2012;3(2).

21. Guilfoyle R, Demers A, Bredeson C, Richardson E, et al. Performance status, but not the hematopoietic cell transplantation comorbidity index (HCT-CI), predicts mortality at a Canadian transplant center. Bone Marrow Transplant. 2009;43(2):133-9.

22. Nakaya A, Mori T, Tanaka M, Tomita M, et al. Does the hematopoietic cell transplantation specific comorbidity index (HCT-CI) predict transplantation outcomes? A prospective multicenter validation study of the Kanto Study Group for Cell Therapy. Biol Blood Marrow Transplant. 2014;20(10):1553-9.

23. Sorror M, Storer B, Sandmaier B, Maloney D, Chauncey T, et al. Hematopoietic cell transplantation-comorbidity index and Karnofsky performance status are independent predictors of morbidity and mortality after allogeneic nonmyeloablative hematopoietic cell transplantation. Cancer. 2008;112(9):1992-2001.

24. Vaughn J, Storer B, Armand P, Raimondi R, Gibson C, et al. Design and Validation of an Augmented Hematopoietic Cell Transplantation-Comorbidity Index Comprising Pretransplant Ferritin, Albumin, and Platelet Count for Prediction of Outcomes after Allogeneic Transplantation. Biol Blood Marrow Transplant. 2015;pii: S1083-8791(15)00231-1. doi: 10.1016/j.bbmt.2015.04.002.

25. Kroger N., Brand R., Niederwieser D., Platzbecker U., Hubel K., Weber T., et al. RIC vs. MAC followed by allogeneic stem cell transplantation for patients with MDS or secondary AML: A prospective, randomized phase III study of the CMWP of the EBMT (RICMAC-Trial). Bone Marrow Transplantation. 2015;50(S5):O006.

26. Algarotti A., Mico C., Corradini P., Falda M., Alessandrino E.P., Fanin R., et al. In vivo T cell depletion with antithymocyte globulin or alemtuzumab for unrelated donor stem cell transplantation with reduced intensity conditioning: Results of a multicenter randomized phase II clinical trial (the global study) from the Gruppo Italiano Trapianto di Midollo Osseo (GITMO). Haematologica. 2013;98(S3):S60.

27. Baron F, Zachee P, Maertens J, Kerre T, Ory A, Seidel L, et al. Non-myeloablative allogeneic hematopoietic cell transplantation following fludarabine plus 2 Gy TBI or ATG plus 8 Gy TLI: a phase II randomized study from the Belgian Hematological Society. J Hematol Oncol. 2015;8:4.

28. Beguin Y., Zachee P., Maertens J., Kerre T., Ory A., Seidel L., et al. Nonmyeloablative Allogeneic Hematopoietic Cell Transplantation Following Fludarabine Plus 2 Gy TBI or ATG Plus 8 Gy TLI: A Phase II Randomized Study from the Belgian Hematological Society. Blood. 2014;124:542-.

29. Blaise D, Maraninchi D, Archimbaud E, Reiffers J, Devergie A, Jouet JP, et al. Allogeneic bone marrow transplantation for acute myeloid leukemia in first remission: a randomized trial of a busulfan-Cytoxan versus Cytoxan-total body irradiation as preparative regimen: a report from the Group d'Etudes de la Greffe de Moelle Osseuse. Blood. 1992;79(10):2578-82.

30. Blaise D, Maraninchi D, Michallet M, Reiffers J, Jouet JP, Milpied N, et al. Long-term follow-up of a randomized trial comparing the combination of cyclophosphamide with total body irradiation or busulfan as conditioning regimen for patients receiving HLA-identical marrow grafts for acute myeloblastic leukemia in first complete remission. Blood. 2001;97(11):3669-71.

31. Blaise D, Tabrizi R, Boher JM, Le Corroller-Soriano AG, Bay JO, Fegueux N, et al. Randomized study of 2 reduced-intensity conditioning strategies for human leukocyte antigen-matched, related allogeneic peripheral blood stem cell transplantation: prospective clinical and socioeconomic evaluation. Cancer. 2013;119(3):602-11.

32. Blume KG, Kopecky KJ, Henslee-Downey JP, Forman SJ, Stiff PJ, LeMaistre CF, et al. A prospective randomized comparison of total body irradiation-etoposide versus busulfan-cyclophosphamide as preparatory regimens for bone marrow transplantation in patients with leukemia who were not in first remission: a Southwest Oncology Group study. Blood. 1993;81(8):2187-93.

33. Bornhauser M., Kienast J., Trenschel R., Burchert A., Hegenbart U., Stadler M., et al. An intergroup randomised trial of standard intensity versus reduced intensity TBI-based conditioning in patients with acute myeloid leukemia in first complete remission. Blood. 2011;118(21):Abstract 157.

34. Champlin RE, Perez WS, Passweg JR, Klein JP, Camitta BM, Gluckman E, et al. Bone marrow transplantation for severe aplastic anaemia: a randomized controlled study of conditioning regimens. Blood. 2007;109(10):4582-5.

35. Clift RA, Buckner CD, Thomas ED, Bensinger WI, Bowden R, Bryant E, et al. Marrow transplantation for chronic myeloid leukemia: a randomized study comparing cyclophosphamide and total body irradiation with busulfan and cyclophosphamide. Blood. 1994;84(6):2036-43.

36. Clift RA, Radich J, Appelbaum FR, Martin P, Flowers ME, Deeg HJ, et al. Long-term follow-up of a randomized study comparing cyclophosphamide and total body irradiation with busulfan and cyclophosphamide for patients receiving allogenic marrow transplants during chronic phase of chronic myeloid leukemia. Blood. 1999;94(11):3960-2.

37. Devergie A, Blaise D, Attal M, Tigaud JD, Jouet JP, Vernant JP, et al. Allogeneic bone marrow transplantation for chronic myeloid leukemia in first chronic phase: a randomized trial of busulfan-cytoxan versus cytoxan-total body irradiation as preparative regimen: a report from the French Society of Bone Marrow Graft (SFGM). Blood. 1995;85(8):2263-8.

38. Grassi A., Mico C., Oldani E., Boschini C., Busca A., Benedetto B., et al. Myeloablative, reduced toxicity versus standard conditioning in AML: a randomized clinical trial from Gruppo Italiano Trapianto di Midollo Osseo (GITMO). Bone Marrow Transplant. 2015;50:s2.

39. Gratwohl A, Hermans J, von Biezen A, Arcese W, de Witte T, Debusscher L, et al. No advantage for patients who receive splenic irradiation before bone marrow transplantation for chronic myeloid leukaemia: results of a prospective randomized study. Bone Marrow Transplant. 1992;10(2):147-52.

40. Gratwohl A, Hermans J, Biezen AV, Arcese W, de Witte T, Debusscher L, et al. Splenic irradiation before bone marrow transplantation for chronic myeloid leukemia: update of a prospective randomized study. Leuk Lymphoma. 1993;11 Suppl 1:227-31.

41. Helenglass G, Powles RL, McElwain TJ, Lakhani A, Milan S, Gore M, et al. Melphalan and total body irradiation (TBI) versus cyclophosphamide and TBI as conditioning for allogeneic matched sibling bone marrow transplants for acute myeloblastic leukaemia in first remission. Bone Marrow Transplant. 1988;3(1):21-9.

42. Kim H., Park J.-H., Lee J.-H., Joo Y.D., Lee W.-S., Bae S.-H., et al. Prospective Randomized Comparison of Cyclophosphamide Versus Cyclophosphamide Plus Fludarabine In Addition to Anti-Thymocyte Globulin for the Conditioning Therapy In Allogeneic Hematopoietic Cell Transplantation for Bone Marrow Failure Syndrome Blood. 2010;116:Abstract 34.

43. Kim H, Lee JH, Joo YD, Bae SH, Hyun MS, Lee JH, et al. A randomized comparison of cyclophosphamide vs. reduced dose cyclophosphamide plus fludarabine for allogeneic hematopoietic cell transplantation in patients with aplastic anaemia and hypoplastic myelodysplastic syndrome. Ann Hematol. 2012;91(9):1459-69.

44. Kornblit B, Maloney DG, Storb R, Storek J, Hari P, Vucinic V, 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. 2013;19(9):1340-7.

45. Lee JH, Joo YD, Kim H, Ryoo HM, Kim MK, Lee GW, et al. Randomized trial of myeloablative conditioning regimens: busulfan plus cyclophosphamide versus busulfan plus fludarabine. J Clin Oncol. 2013;31(6):701-9.

46. Liu DH, Xu LP, Zhang XH, Wang Y, Yan CH, Wang JZ, et al. Substitution of cyclophosphamide in the modified BuCy regimen with fludarabine is associated with increased incidence of severe pneumonia: a prospective, randomized study. Int J Hematol. 2013;98(6):708-15.

47. Liu H, Zhai X, Song Z, Sun J, Xiao Y, Nie D, 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 multicenter study. J Hematol Oncol. 2013;6:15.

48. Rambaldi A., Grassi A., Mico M.C., Busca A., Bruno B., Cavattoni I., et al. Randomized Trial of Busulfan with Cyclophosphamide Versus Busulfan with Fludarabine As Preparative Regimen to Allogeneic Hematopoietic Stem Cell Transplantation in Patients with Acute Myeloid Leukemia: A Study from the Gruppo Italiano Trapianto Midollo Osseo (GITMO). Blood. 2014;124(21):727-.

49. Ringden O, Ruutu T, Remberger M, Nikoskelainen J, Volin L, Vindelov L, et al. A randomized trial comparing busulfan with total body irradiation as conditioning in allogeneic marrow transplant recipients

with leukemia: a report from the Nordic Bone Marrow Transplantation Group. Blood. 1994;83(9):2723-30.

50. Ringden O, Ruutu T, Remberger M, Nikoskelainen J, Volin L, Vindelov L, et al. A randomized trial comparing busulfan vs total body irradiation in allogeneic marrow transplant recipients with hematological malignancies. Transplant Proc. 1994;26(3):1831-2.

51. Ringden O, Remberger M, Ruutu T, Nikoskelainen J, Volin L, Vindelov L, et al. Increased risk of chronic graft-versus-host disease, obstructive bronchiolitis, and alopecia with busulfan versus total body irradiation: long-term results of a randomized trial in allogeneic marrow recipients with leukemia. Nordic Bone Marrow Transplantation Group. Blood. 1999;93(7):2196-201.

52. Algarotti A., Mico C., Corradini P., Falda M., Alessandrino E.P., Fanin R., et al. In vivo T cell depletion with antithymocyte globulin or alemtuzumab for unrelated donor stem cell transplantation with reduced intensity conditioning: Results of a multicenter randomized phase II clinical trial (the global study) from the Gruppo Italiano Trapianto di Midollo Osseo (GITMO). Haematologica. 2013;98(s3):17.

4. REPORT APPENDICES

Cite this report as: Wolfe D, Yazdi F, Hutton B, Moher D, Bredeson C, Kekre N, Allan D. Comparison of Conditioning Regimens for Patients Undergoing Hematopoietic Stem Cell Transplantion: A Systematic Review and Meta-analysis (2017).

Appendix 1. Literature Search Strategies

Appendix 2. Additional Characteristics of Included Studies

Appendix 3. Risk of Bias Assessments of Included Studies

Appendix 4. Summary of Results from Secondary Analyses

Appendix 5. Summary of Results for Subgroups Reported for All Outcomes

Appendix 6. Summary of Findings from Random Effects NMAs for the Review

Appendix 7. PRISMA NMA Checklist

Appendix 1: Literature search strategies

Database: Ovid MEDLINE(R) In-Process & Other Non-Indexed Citations and Ovid MEDLINE(R) <1946 to Present> **June 12, 2013; updated June 2015**

---

1 Hematopoietic Stem Cell Transplantation/
2 ((h?ematopoietic adj3 transplant$) or hsct$).tw.
3 peripheral blood cell transplant$.tw.
4 peripheral blood stem cell transplant$.tw.
5 Bone Marrow Transplantation/
6 bone marrow transplant$.tw.
7 STEM CELL TRANSPLANTATION/
8 stem cell transplant$.tw.
9 stem cell therap$.tw.
10 PERIPHERAL BLOOD STEM CELL TRANSPLANTATION/
11 peripheral stem cell transplant$.tw.
12 H?ematopoietic peripheral blood stem cell transplant$.tw.
13 or/1-12
14 randomized controlled trial.pt.
15 controlled clinical trial.pt.
16 randomized.ab.
17 placebo.ab.
18 clinical trials as topic.sh.
19 randomly.ab.
20 trial.ti.
21 or/14-20
22 exp animals/ not humans.sh.
23 21 not 22
24 13 and 23

Appendix 2: Additional characteristics of included studies

Table 15: Distribution of underlying disease across studies of regimens for GVHD prophylaxis

Underlying disease	# of studies that included any patients with the underlying disease (18 studies overall)	Number of patients with the underlying disease (2,355 patients overall)
ALL	6	170
AML	13	1,100
CML	9	617
Non-Hodgkin lymphoma	3	79
Myelodysplastic syndrome	6	57
Multiple myeloma	3	68
Other malignant disease[a]	3	45
Any acute leukemia[b]	11	1,270
Any lymphoma[c]	4	83
Aplastic anaemia	2	209
Other non-malignant disease[d]	2	7

[a] includes Hodgkin's lymphoma, chronic lymphocytic leukemia and Waldenstrom disease
[b] includes the sum of ALL and AML patients listed separately in table
[c] includes non-Hodgkin lymphoma patients listed separately in table as well as patients reported as "lymphoma"
[d] includes primary myelofibrosis and myeloproliferative disorder

Appendix 3: Risk of bias assessments of included studies

Risk of bias assessments of included studies										
	Selection bias		Performance bias	Detection bias	Attrition bias	Reporting bias	Other bias	Summary assessment of bias by outcome[b]		
Study	Random sequence generation	Allocation concealment	Blinding of patients & personnel	Blinding of outcome assessor	Incomplete outcome data	Selective outcome reporting	Other bias[a]	Mortality and Relapse	GVHD	Harms[c]
Baron (2015)[27, 28]										
Rambaldi (2014)[38, 48]										
Blaise (2013)[31]										
Kornblit (2013)[44]										
Lee (2013)[45]										
Liu (2013)[47]										
Liu (2013)[46]										
Algarotti[26, 52]										
Kim (2012)[42, 43]										
Bornhauser (2011)[33]										
Champlin (2007)[34]										
Ringdén (1999)[49-51]										
Devergie (1995)[37]										
Clift (1994)[35, 36]										
Blume (1993)[32]										
Gratwhol (1993)[39, 40]										
Blaise (1992, 2001)[29, 30]										
Helenglass (1988)[41]										

Green = low ROB; yellow = unclear ROB; red = high ROB; blue = outcome not reported

[a]Studies with treatment groups that were unbalanced with respect to patient demographics were considered to have a high risk of "other" bias.

[b]Summary assessments were based on the highest risk identified in any of the key domains for each outcome group. If all key domains demonstrated low risk, the summary assessment was low ROB. If one or more of the key domains demonstrated unclear risk and the rest were low risk, the summary assessment was unclear ROB. If one or more of the key domains demonstrated high risk, the summary assessment was high risk. For mortality outcomes, which were considered objective, performance bias, detection bias, and industry funding not considered to be key domains and were excluded from the summary assessments. For the subjective outcomes of efficacy and harms, all domains were considered key.

[c]Harms included veno-occlusive disease, bronchiolitis obliterans, infection, and other harms

Appendix 4: Findings from Secondary Analyses

As described in the review's main text, during the conduct of this review several disjoint networks were encountered which required separate analysis in addition to those networks that were chosen as the primary focus for the different endpoints based upon input from clinical experts. Additionally, in some cases there was a need to provide a narrative overview of study findings because meta-analyses were not feasible. The sections below provide all of this additional information with the objective of providing all findings encountered during the review process.

Findings: Overall mortality—secondary network meta-analysis

Three studies[27, 31, 44], evaluating a total of 318 patients formed a secondary network that was disjoint from the primary network that was presented in the main body of this report. Four conditioning regimens were evaluated in these studies.

Network diagram for the secondary NMA of overall mortality

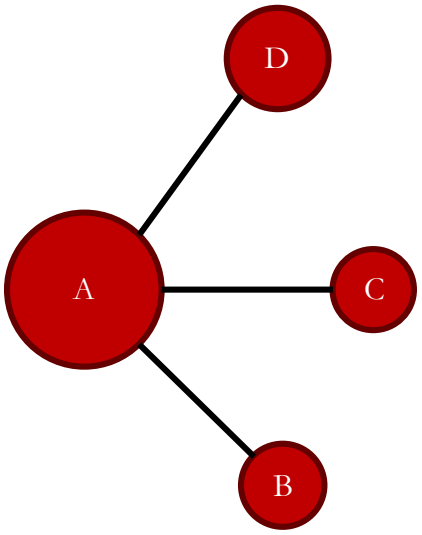


Figure legend
A = FLUD+TBI
B = ATG+TLI
C = TBI
D = ATG+FLUD+BU

ATG = anti-thymocyte globulin
BU = busulfan
FLUD = fludarabine
TBI = total body irradiation
TLI = total lymphoid irradiation

Results from traditional pairwise meta-analyses

For follow-up times of 100-days, 1-year, and 5-years, the table below presents summaries of pairwise estimates of overall mortality that were derived from direct evidence (i.e., the head-to-head trials), summarized alongside the numbers of available trials and patients for each direct comparison, as well as the related summary estimate derived from the network meta-analysis. All direct comparisons at all time points were informed by single studies. None of the comparisons demonstrated significant differences in risk of mortality on pairwise meta-analysis. The estimates from pairwise meta-analyses compared well with those derived from network meta-analysis.

Summary of results from pairwise meta-analysis and NMA, 100-day overall mortality				
Comparison		**# of Trials (patients)**	**Direct Estimate OR (95% CrI)**	**FE NMA Estimate OR (95% CI)**
			***Values <1 favor comparator**	
Comparator	**Reference**			
100-day overall mortality				
ATG+TLI	FLUD+TBI	1 (94)	0.80 (0.17–3.81)	0.79 (0.14–4.04)
FLUD+TBI	ATG+FLUD+BU	1 (139)	0.54 (0.15–1.92)	0.52 (0.12–1.83)
FLUD+TBI	TBI	1 (85)	0.51 (0.09–2.96)	0.47 (0.05–2.74)
1-year overall mortality				
ATG+FLUD+BU	FLUD+TBI	1 (139)	0.94 (0.44–2.03)	0.94 (0.43–2.04)
FLUD+TBI	ATG+TLI	1 (94)	0.55 (0.21–1.46)	0.54 (0.20–1.43)
FLUD+TBI	TBI	1 (85)	0.46 (0.15–1.36)	0.44 (0.13–1.31)
5-year overall mortality				
ATG+TLI	FLUD+TBI	1 (94)	0.98 (0.44–2.22)	0.98 (0.43–2.27)
FLUD+TBI	ATG+FLUD+BU	1 (139)	0.97 (0.49–1.90)	0.97 (0.49–1.90)
FLUD+TBI	TBI	1 (85)	0.48 (0.20–1.15)	0.47 (0.19–1.13)
ATG = anti-thymocyte globulin; BU = busulfan; FLUD = fludarabine; TBI = total body irradiation; TLI = total lymphoid irradiation				

Results from network meta-analysis

For all follow-up periods, 3 of the 6 possible pairwise comparisons were informed only by indirect evidence. The remaining 3 comparisons were informed by head-to-head trials; however, these direct comparisons were all comprised of single studies with limited numbers of patients. Model fit statistics for all follow-up periods indicated that both the FE and RE models had an adequate fit. All NMA and pairwise results presented for the overall mortality outcome are derived from FE models.

Model fit for 100-day, 1- and 5-year mortality outcomes from FE and RE NMA							
Follow-up time	**# of comparisons informed only by indirect evidence of total possible comparisons**	**# of single-study direct comparisons**	**Number of data points**	**FE total deviance residual**	**RE total deviance residual**	**FE DIC**	**RE DIC**
100-day	3of 6	3 of 3	6	6.313	6.315	30.997	31.006
1-year	3of 6	3 of 3	6	6.139	6.074	36.504	36.373
5-year	3of 6	3 of 3	6	6.078	6.092	38.431	38.457
DIC = deviance information criterion; FE = fixed effects mode; RE = random effects model							

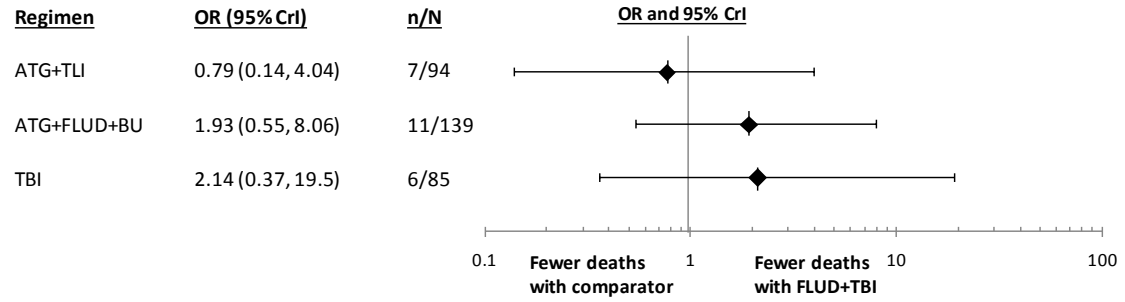
Comparisons versus FLUD+TBI

The figure below presents forest plots summarizing comparisons of all conditioning regimens in the evidence network to the chosen reference regimen, FLUD+TBI, at the follow-up times of interest. All of

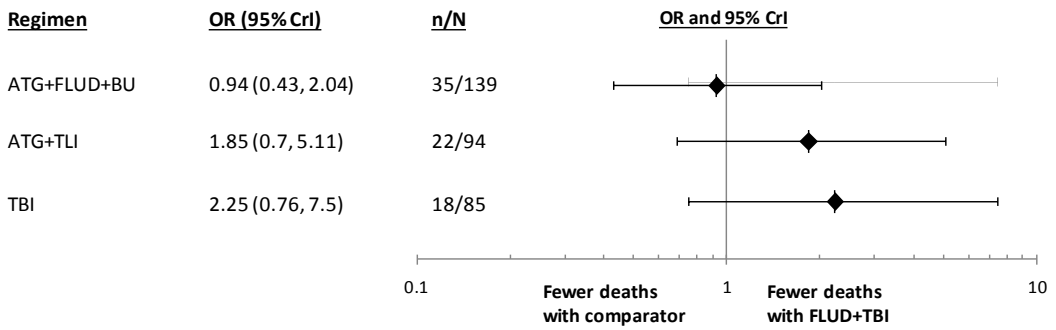
the interventions were associated with credible intervals that included 1, indicating they were not significantly different from FLUD+TBI in reducing overall mortality at the selected follow-up time.

Forest plots of comparisons versus FLUD+TBI, Overall mortality in secondary network, all follow-up times

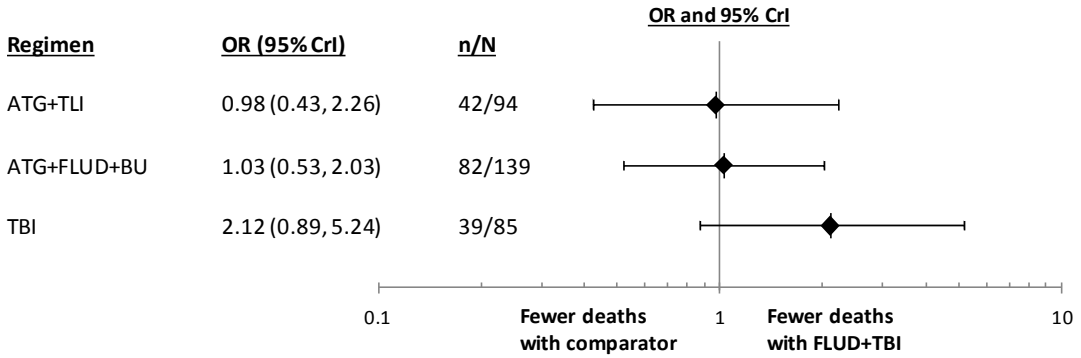
100-day overall mortality



1-year overall mortality



5-year overall mortality



Odds ratios from network meta-analysis comparing each regimen to the reference treatment of FLUD+TBI are presented along with the number of events and patients per regimen. Odds ratios <1 suggest more risk of mortality with FLUD+TBI while odds ratios >1 suggest less risk of mortality with FLUD+TBI.

Comparisons between all conditioning regimens

The figure below presents league tables of the estimates for all 6 pairwise comparisons generated by NMA for each of the follow-up times of interest. Some of the comparisons included in the league tables should be interpreted carefully, specifically those which are not informed by head to head trials or where the comparison was based on compound indirect evidence.

League table summary of pairwise comparisons in the network meta-analysis of the secondary network for 100-day, 1-, and 5-year overall mortality

100-day overall mortality

0.74			
ATG+TLI	**0.68**		
0.79 (0.14 – 4.04)	**FLUD+TBI**	**0.30**	
0.40 (0.04 – 3.19)	0.52 (0.12 – 1.83)	**ATG+FLUD+BU**	**0.29**
0.35 (0.02 – 4.12)	0.47 (0.05 – 2.74)	0.90 (0.07 – 8.61)	**TBI**

1-year overall mortality

0.77			
ATG+FLUD+BU	**0.75**		
0.94 (0.43 – 2.04)	**FLUD+TBI**	**0.29**	
0.51 (0.14 – 1.76)	0.54 (0.20 – 1.43)	**ATG+TLI**	**0.19**
0.42 (0.10 – 1.58)	0.44 (0.13 – 1.31)	0.81 (0.18 – 3.63)	**TBI**

5-year overall mortality

0.66			
FLUD+TBI	**0.65**		
1.02 (0.44 – 2.33)	**ATG+TLI**	**0.61**	
0.97 (0.49 – 1.90)	0.94 (0.32 – 2.80)	**ATG+FLUD+BU**	**0.08**
0.47 (0.19 – 1.13)	0.46 (0.14 – 1.57)	0.49 (0.16 – 1.48)	**TBI**

LEGEND

ATG = anti-thymocyte globulin; BU = busulfan; FLUD = fludarabine; TBI = total body irradiation; TLI = total lymphoid irradiation

1 link **Direct evidence**	**2 links** **Simple indirect**

Findings: Acute GVHD—narrative summary of pairwise comparisons not included in the network meta-analysis

Three pairs of comparisons were disjoint from the network and could not be included in the network meta-analysis. The three studies that evaluated these comparisons have been summarized with their findings in the table below.

Two of the studies demonstrated significant differences in the risk of aGVHD in the conditioning regimens that they compared. Liu et al.[46] compared BU+FLUD and BU+CY regimens that had been modified at their institution to include hydroxyurea, semustine, and cytosine arabinoside. Part way through the trial, a significantly increased risk of severe pneumonia was found in the mBU+FLUD arm and the trial was suspended. The findings below also demonstrate a significantly increased risk of aGVHD in the mBU+FLUD arm.

Algarotti et al.[26, 52] found a significantly increased risk of "> Grade II" aGVHD in patients receiving ATG+THIO+MELPH+CY compared to those receiving ALZ+MELPH+FLUD. This study was only published in abstract form, with little supporting information, except that the study was a prospective Phase II clinical trial conducted following evidence of tolerability and effectiveness of both regimens in preliminary research. There was no evidence in the RCT literature of other institutions using these specific conditioning regimens.

Gratwhol et al.[40] evaluated the effects of adding total lymphoid irradiation to the standard conditioning regimens at a variety of institutions. The standard conditioning regimens varied by study site. The risk of aGVHD did not change significantly when TLI was added.

Summary of results: Risk of aGVHD in comparisons that were disjoint from the main treatment network

Author	Follow-up	Outcome definition	Treatment regimen	Group risk	OR (95% CI)
Liu (2013)[46]	100 days	NR	mBU+FLUD	18/52 (35%)*	**2.59 (1.03–6.47)**
			mBU+CY	9/53 (17%)	
Algarotti (2013)[26, 52]	100 days	"> Grade II aGVHD"	ALZ+MELPH+FLUD	26/58 (45%)*	**2.67 (1.24–5.80)**
			ATG+THIO+MELPH+CY	37/54 (69%)	
Gratwhol (1993)[39, 40]	100 days	NR	Standard conditioning regimens[a]	35/114 (31%)	1.10 (0.62–1.94)
			TLI+standard conditioning regimens	33/115 (29%)	

*denotes the reference group for each pairwise comparison
[a]Standard conditioning regimens varied by study site
ALZ = alemtuzumab; ATG = anti-thymocyte globulin; BU = busulfan; CY = cyclophosphamide; FLUD = fludarabine; mBU+CY = BU+CY+ hydroxyurea+semustine+cytosine arabinoside; mBU+FLUD = BU+FLUD+hydroxyurea+semustine+cytosine arabinoside; MELPH = melphalan; THIO = thiotepa

Findings: Extensive chronic GVHD in evaluable patients—secondary network meta-analysis and narrative summary of pairwise comparison not included in the primary or secondary networks

Secondary network: Results from traditional pairwise meta-analysis

For the secondary network evaluating extensive cGVHD in patients alive at 100 days, two studies[27, 31] were included. The table below presents summaries of pairwise estimates that were derived from direct evidence (i.e., the head-to-head trials), alongside the numbers of available trials and patients for each direct comparison, as well as the related summary estimate derived from the network meta-analysis.

Secondary network diagram for the NMA of extensive cGVHD in evaluable patients

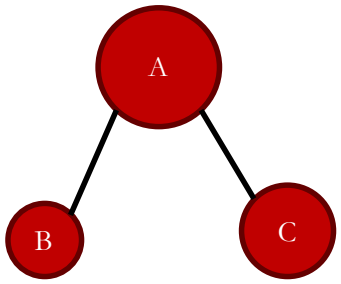


Figure legend
A = FLUD+TBI
B = ATG+TLI
C = ATG+BU+FLUD

ATG = anti-thymocyte globulin
BU = busulfan
FLUD = fludarabine
TBI = total body irradiation
TLI = total lymphoid irradiation

The two comparisons with direct evidence had 1 study each informing them. Pairwise meta-analysis suggested there was a significant reduction of risk of extensive cGVHD in patients alive at 100 days post-transplant for ATG+TLI compared to FLUD+TBI but not for ATG+BU+FLUD compared to FLUD+TBI. Estimates from traditional pairwise meta-analyses compared well with estimates derived from network meta-analysis.

Summary of results from meta-analysis and NMA of secondary network, Extensive cGVHD in evaluable patients

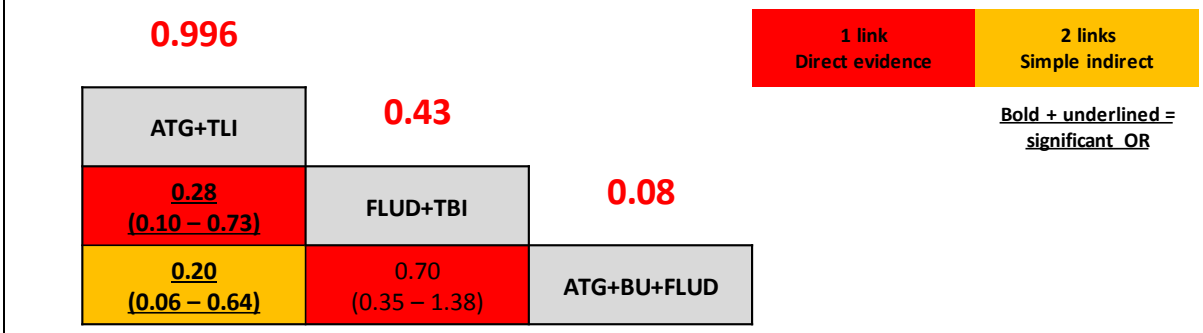
Comparison		# of Trials (patients)	Direct Estimate OR (95% CrI)	NMA Estimate OR (95% CI)
			**Values <1 favor comparator*	
Comparator	**Reference**			
ATG+TLI	FLUD+TBI	1 (87)	**0.29 (0.11–0.78)**	**0.28 (0.10–0.73)**
FLUD+TBI	ATG+BU+FLUD	1 (134)	0.70 (0.35–1.39)	0.70 (0.35–1.38)

Secondary network: Results from network meta-analysis

Two of the three possible comparisons in the secondary network were informed by direct evidence, both of which were informed by 1 study each. Model fit statistics for the NMA indicated that both the fixed-effects (FE) and random-effects (RE) models had an adequate fit. Posterior residual deviance values for the FE and RE models of 4.07 and 4.05, respectively, were obtained, both of which demonstrated a desirable fit, given the 4 data points in the model. DIC values (25.391 and 25.354, respectively) suggested similar fit between both RE and FE models.

The figure below presents a league table of the estimates for all 3 pairwise comparisons generated by NMA. ATG+TLI was associated with a significantly reduced risk of extensive cGVHD in patients alive at 100 days compared to FLUD+TBI and ATG+BU+FLUD.

League table summary of pairwise comparisons from NMA of secondary network, Odds Ratios with 95% CrI, Extensive cGVHD in evaluable patients



Interventions are sorted from left to right in order of decreasing SUCRA value (i.e., preferred treatments appear first). For each comparison shown, the upper/left-most regimen is the comparator group while the lower/right-most treatment is the reference treatment; a value <1 suggests fewer cases of extensive chronic GVHD with the comparator than with the reference group. Statistically significant differences (i.e., estimates with a 95% credible interval excluding 1) are bolded and underlined. Comparisons with direct evidence and simple indirect evidence have been highlighted in red and orange, respectively.

Narrative summary of pairwise comparison not included in the primary or secondary networks for extensive cGVHD in evaluable patients

One pair of interventions was disjoint from both the primary and secondary networks and could not be included in the network meta-analyses. The single study[46] evaluating this comparison has been summarized with its findings in the table below.

No significant difference in the risk of extensive cGVHD was found in the conditioning regimens compared. The authors of this paper used BU+FLUD and BU+CY regimens that had been modified at their institution to include hydroxyurea, semustine, and cytosine arabinoside. Part way through the trial, a significantly increased risk of severe pneumonia was found in the mBU+FLUD arm and the trial was suspended.

Summary of results: Risk of aGVHD in comparisons that were disjoint from the main treatment network

Author	Follow-up	Outcome definition	Treatment regimen	Group risk	OR (95% CI)
Liu (2013)[46]	3.3 years	Extensive cGVHD	mBU+FLUD	9/49 (18%)	1.98 (0.61–6.41)
			mBU+CY	5/49 (10%)*	

*denotes the reference group for the pairwise comparison
BU = busulfan; CY = cyclophosphamide; FLUD = fludarabine; mBU+CY = BU+CY+ hydroxyurea+semustine+cytosine arabinoside; mBU+FLUD = BU+FLUD+hydroxyurea+semustine+cytosine arabinoside;

Findings: Overall cGVHD in evaluable patients—network meta-analysis and narrative summary of pairwise comparisons not included in network

Seven studies[26, 29, 32, 44, 46, 47, 51] evaluating 707 patients alive at 100 days had data available for the overall cGVHD outcome. Ten conditioning regimens were evaluated in these studies; however, six of these interventions formed three disjoint pairs separate from the primary network of four regimens evaluated in four studies[29, 32, 47, 51]. The three disjoint pairs have been summarized narratively after the primary network meta-analysis below.

Network diagrams for the NMA of overall cGVHD in evaluable patients

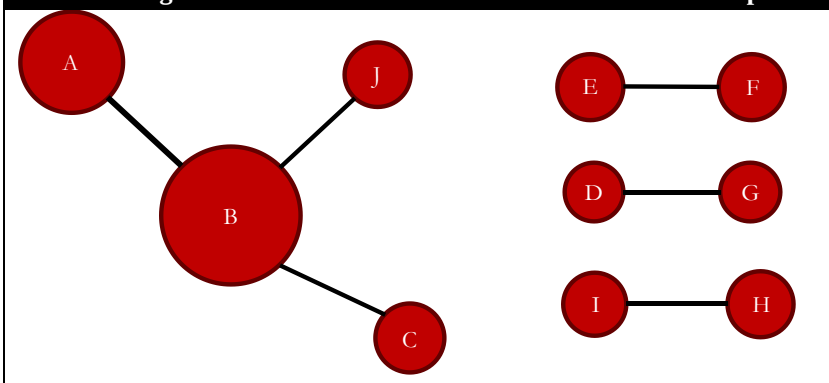


Figure legend
A = CY+TBI
B = BU+CY
C = BU+FLUD
D = FLUD+TBI
E = mBU+FLUD
F = mBU+CY
G = TBI
H = ALZ+MELPH+FLUD
I = ATG+THIO+MELPH+CY
J = ETP+TBI

ATG = anti-thymocyte globulin; ALZ = alemtuzumab; BU = busulfan; CY = cyclophosphamide; FLUD = fludarabine; ETP = etoposide; mBU+CY = BU+CY+ hydroxyurea+semustine+cytosine arabinoside; mBU+FLUD = BU+FLUD+hydroxyurea +semustine+cytosine arabinoside; MELPH = melphalan; TBI = total body irradiation; THIO = thiotepa; TLI = total lymphoid irradiation

Primary network: Results from traditional pairwise meta-analysis

For the network evaluating overall cGVHD in patients alive at 100 days, the table below presents summaries of pairwise estimates that were derived from direct evidence alongside the numbers of available trials and patients for each direct comparison, as well as the related summary estimate derived from the network meta-analysis. Four studies were included in this analysis[29, 32, 47, 51].

One of the three comparisons with direct evidence had 2 studies informing it (n = 239 patients), while the other two had one study each of approximately 100 patients. Pairwise meta-analyses did not demonstrate significant differences between any of the interventions compared. Estimates from traditional pairwise meta-analyses compared well with estimates derived from network meta-analysis.

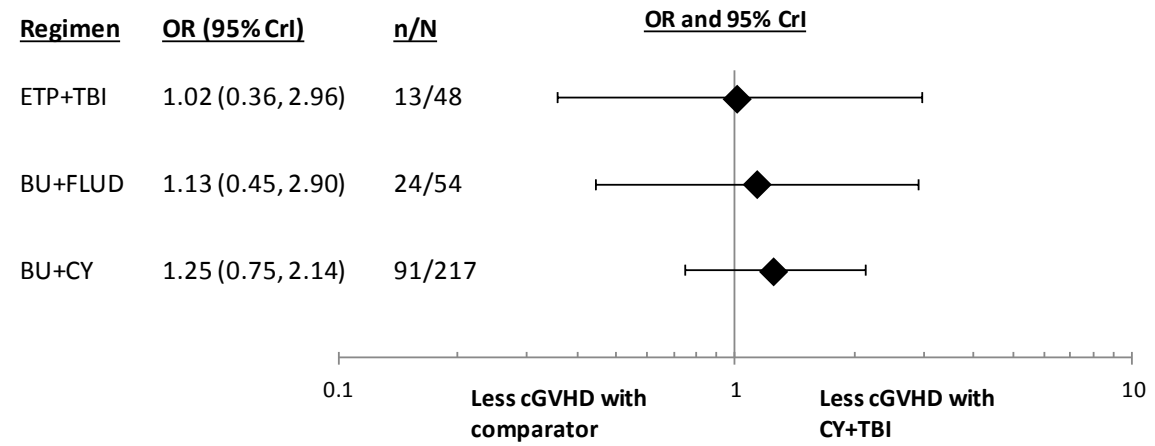
Summary of results from meta-analysis and NMA of secondary network, Overall cGVHD in evaluable patients					
Comparison		# of Trials (patients)	Heterogeneity (I^2)	Direct Estimate OR (95% CrI)	NMA Estimate OR (95% CI)
				*Values <1 favor comparator	
Comparator	Reference				
CY+TBI	BU+CY	2 (239)	57.834	0.80 (0.47–1.35)	0.80 (0.47–1.34)
ETP+TBI	BU+CY	1 (93)	—	0.82 (0.34–2.02)	0.81 (0.32–2.03)
BU+FLUD	BU+CY	1 (105)	—	0.90 (0.42–1.94)	0.91 (0.42–1.97)

Primary network: Results from network meta-analysis

Three of the six comparisons were informed by direct evidence, and two of these three direct comparisons were informed only by single studies, with limited numbers of patients. Model fit statistics for the NMA indicated that both the fixed-effects (FE) and random-effects (RE) models had an adequate fit. Posterior residual deviance values for the FE and RE models of 9.501 and 8.246, respectively, were obtained, both of which demonstrated a desirable fit, given the 8 data points in the model. DIC values (51.422 and 51.087, respectively) suggested similar fit; a FE model was preferred due to the number of single-study connections in the network.

The figure below presents a forest plot summarizing comparisons of all conditioning regimens in the evidence network to the chosen reference therapy, CY+TBI. All regimens were associated with credible intervals that included 1 and, thus, were not significantly different from CY+TBI in reducing the risk of overall cGVHD.

Forest plot of comparisons from NMA versus CY+TBI—Overall cGVHD in evaluable patients

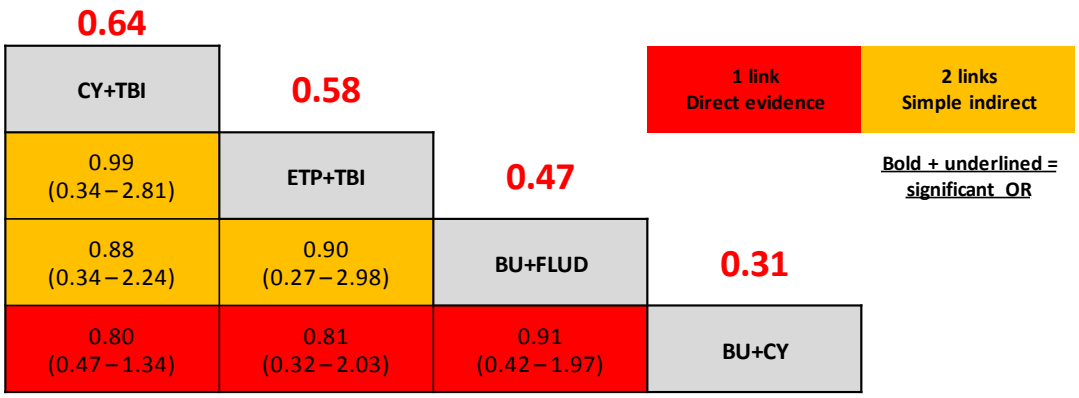


Odds ratios from network meta-analysis comparing each regimen to the reference regimen of CY+TBI are presented along with the number of events and patients per regimen. Odds ratios <1 suggest more risk of overall cGVHD with CY+TBI while odds ratios >1 suggest less risk of acute GVHD with CY+TBI.

The figure shown below presents a league table of the estimates for all comparisons generated by NMA. None of the conditioning regimens was associated with a significantly reduced risk of extensive cGVHD

in patients alive at 100 days compared to any of the other regimens. CY+TBI was ranked highest with respect to efficacy in reducing the risk of overall cGVHD; however, there was little difference in the SUCRA values of the top-ranked treatments, indicating their relative rankings were uncertain.

League table summary of pairwise comparisons from NMA of primary network, Odds Ratios with 95% CrI, Overall cGVHD in evaluable patients



Interventions are sorted from left to right in order of decreasing SUCRA value (i.e., preferred treatments appear first). For each comparison shown, the upper/left-most regimen is the comparator group while the lower/right-most treatment is the reference treatment; a value <1 suggests fewer cases of overall chronic GVHD with the comparator than with the reference group. Statistically significant differences (i.e., estimates with a 95% credible interval excluding 1) are bolded and underlined. Comparisons with direct evidence and simple indirect evidence have been highlighted in red and orange, respectively.

Narrative summary of pairwise comparisons not included in the primary network for overall cGVHD in evaluable patients

Three pairs of interventions were disjoint from the primary networks and could not be included in the network meta-analysis. The three studies[26, 44, 46] evaluating these comparisons have been summarized with their findings in the table below.

In one of the three studies[44], there appeared to be significant difference in the risk of overall cGVHD between FLUD+TBI and TBI alone, when we compared the frequency data using an odds ratio. However, a survival analysis conducted by the study authors demonstrated no significant difference in the hazard of overall cGVHD between the two regimens (HR = 1.52; 95% CI: 0.9–2.7; p = 0.14).

Liu et al.[46] compared BU+FLUD and BU+CY regimens that had been modified at their institution to include hydroxyurea, semustine, and cytosine arabinoside. Part way through the trial, a significantly increased risk of severe pneumonia was found in the mBU+FLUD arm and the trial was suspended. No significant difference in the risk of overall cGVHD was found between the regimen arms.

Algarotti et al.[26] found no significant difference in the risk of overall cGVHD in patients receiving ATG+THIO+MELPH+CY compared to those receiving ALZ+MELPH+FLUD. This study was a prospective Phase II clinical trial conducted following evidence of tolerability and effectiveness of both regimens in preliminary research. There was no evidence in the RCT literature of other institutions using these specific conditioning regimens.

Summary of results: Risk of overall cGVHD in comparisons that were disjoint from the primary treatment network					
Author	**Follow-up**	**Outcome definition**	**Treatment regimen**	**Group risk**	**OR (95% CI)**
Liu (2013)[46]	3.3 years	Overall cGVHD	mBU+FLUD	20/49 (41%)	1.19 (0.53–2.68)
			mBU+CY	18/49 (37%)*	
Algarotti (2013)[26, 52]	1.3 (0.8–3.3) years	Grading not reported	ALZ+MELPH+FLUD	23/48 (48%)*	1.04 (0.46–2.35)
			ATG+THIO+MELPH+CY	22/45 (49%)	
Kornblit (2013)[44]	3 years	Grading not reported	FLUD+TBI	30/39 (77%)	**3.02 (1.14–7.95) (Non-significant in survival analysis reported by authors)**
			TBI	21/40 (53%)*	
*denotes the reference group for each pairwise comparison ALZ = alemtuzumab; ATG = anti-thymocyte globulin; BU = busulfan; CY = cyclophosphamide; FLUD = fludarabine; mBU+CY = BU+CY+ hydroxyurea+semustine+cytosine arabinoside; mBU+FLUD = BU+FLUD+hydroxyurea+semustine+cytosine arabinoside; MELPH = melphalan; TBI = total body irradiation; THIO = thiotepa					

Findings, Relapse, 2–3 years follow-up:
narrative summary of pairwise comparisons not included in network

Three pairs of comparisons were disjoint from the network and could not be included in the network meta-analysis. The three studies[26, 44, 46] that evaluated these comparisons have been summarized with their findings in the table below. None of the studies demonstrated significant differences in the risk of relapse in the conditioning regimens that they compared.

Summary of results: Risk of relapse after 2–3 years of follow-up in comparisons that were disjoint from the main treatment network					
Author	**Follow-up**	**Outcome definition**	**Treatment regimen**	**Group risk**	**OR (95% CI)**
Liu (2013)[46]	3.25 years	Recurrence of hematologic leukemia	mBU+FLUD	14/53 (26%)	2.31 (0.85–6.30)
			mBU+CY	7/52 (13%)*	
Algarotti (2013)[26, 52]	2 years	Relapse	ALZ+MELPH+FLUD	14/58 (24%)*	1.57 (0.69–3.59)
			ATG+THIO+MELPH+CY	18/54 (33%)	
Kornblit (2013)[44]	3 years (unclear)	Relapse	FLUD+TBI	16/41 (39%)*	2.26 (0.95–5.38)
			TBI	26/44 (59%)	
*denotes the reference group for each pairwise comparison ALZ = alemtuzumab; ATG = anti-thymocyte globulin; BU = busulfan; CY = cyclophosphamide; FLUD = fludarabine; mBU+CY = BU+CY+ hydroxyurea+semustine+cytosine arabinoside; mBU+FLUD = BU+FLUD+hydroxyurea+semustine+cytosine arabinoside; MELPH = melphalan; TBI = total body irradiation; THIO = thiotepa					

Findings: Veno-occlusive disease—narrative summary of pairwise comparisons not included in the primary network

One pairwise comparison was disjoint from the treatment network and could not be included in the network meta-analysis. The single study[46] that evaluated this comparison has been summarized with its findings in the table below. There were no cases of VOD in either treatment arm after 3.25 years of follow-up.

Summary of results: Risk of VOD in comparisons that were disjoint from the main treatment network					
Author	**Follow-up**	**Outcome definition**	**Treatment regimen**	**Group risk**	**OR (95% CI)**
Liu (2013)[46]	3.25 years	Sinusoidal obstructive syndrome	mBU+FLUD	0/52 (0%)	Not estimable
			mBU+CY	0/53 (0%)	
BU = busulfan; CY = cyclophosphamide; FLUD = fludarabine; mBU+CY = BU+CY+ hydroxyurea+semustine+cytosine arabinoside; mBU+FLUD = BU+FLUD+hydroxyurea+semustine+cytosine arabinoside					

Findings for studies evaluating only patients with aplastic anaemia

Two studies[34, 43] included only patients with aplastic anaemia and were considered too dissimilar to the remaining studies, with respect to patient risk or disease status, to be included in network meta-analyses. Both studies evaluated CY+ATG against differing competing conditioning regimens, allowing for a 3-node network meta-analysis to be conducted, if possible. However, only one outcome was common to both studies (cGVHD) and the required denominator (i.e., the number of patients alive at 100 days of follow-up) was not reported by one of the studies. Thus, all outcomes have been summarized narratively, grouped by study.

Summary of findings for Champlin et al. (2007)[34]

A total of 130 patients were included in the study by Champlin et al.[34] that compared CY to CY+ATG as conditioning regimens for the treatment of aplastic anaemia patients. Patient demographics have been summarized in the table below. Patient age ranged from 1–51 years, with most patients (97%) receiving bone marrow-derived HSCTs. All patients were related to their donors and were fully HLA matched. CsA+MTX was used for GVHD prophylaxis.

Three outcomes were reported by Champlin et al.[34]: 100-day Kaplan-Meier for overall mortality, and 5-year cumulative incidences of acute and chronic GVHD. No significant differences were found between the competing conditioning regimens for any outcomes.

Summary of results for all reported outcomes: Champlin et al. (2007)[34]					
Outcome	**Outcome definition**	**Follow-up time**	**Treatment regimen**	**Findings by treatment arm**	**Study conclusions**
Overall mortality	—	100-day KM estimate	CY	12% of 60 (95% CI: 5–21%)	NSD (p = 0.38)
			CY+ATG	7% of 70 (95% CI: 2–14%)	
aGVHD	Grades II–IV	100-day cumulative incidence	CY	18% of 60 (95% CI: 10–29%)	NSD
			CY+ATG	11% of 70 (95% CI: 6–22%)	
cGVHD	Staging not	5-year	CY	21% of 60	NSD

Summary of results for all reported outcomes: Champlin et al. (2007)[34]					
Outcome	**Outcome definition**	**Follow-up time**	**Treatment regimen**	**Findings by treatment arm**	**Study conclusions**
	reported	cumulative incidence		(95% CI: 12–33%)	
			CY+ATG	32% of 70 (95% CI: 21–44%)	
ATG = anti-thymocyte globulin; CY = cyclophosphamide					

Summary of findings for Kim et al. (2012)[42, 43]

A total of 83 patients were included in the study by Kim et al.[42, 43] that compared CY+ATG to CY+FLUD+ATG as conditioning regimens for the treatment of aplastic anaemia patients. Patient demographics have been summarized earlier in the main text of the report. Patient age ranged from 15–60 years, with most patients (71%) receiving bone marrow-derived HSCTs and 25% receiving HSCTs derived from peripheral blood. The proportion of patients receiving related-donor transplants and the proportion fully HLA matched were not reported. CsA+MTX+steroids was used for GVHD prophylaxis.

The outcomes reported by Kim et al.[42, 43] have been summarized in the table below. Significant differences in the competing conditioning regimens were identified for two harms. The addition of fludarabine to CY+ATG significantly reduced the cumulative incidence of any regimen-related toxicity and of pulmonary complications ($p = 0.003$ and $p = 0.011$, respectively; see table footnote for definitions). Conclusions regarding the relative effects of the regimens on pneumonia/pneumonitis and on CMV reactivation were not reported by the authors; however, our calculations demonstrated there to be no significant difference between the regimens (pneumonia: OR = 3.62; 95% CI = 0.69–19.10; CMV reactivation: OR = 1.30; 95% CI = 0.55–3.12). After subgroup analysis, the significant differences between the regimens for these two outcomes appeared to be driven by donor type, with the addition of FLUD significantly reducing regimen-related toxicity and pulmonary complications in patients receiving matched-sibling donor transplants but not in patients receiving alternative donor transplants.

Summary of results for all reported outcomes: Kim et al. (2012)[42, 43]					
Outcome	**Outcome definition**	**Follow-up time**	**Treatment regimen**	**Findings by treatment arm**	**Study conclusions**
Overall/ Non-relapse mortality	There were no relapses, so OM = NRM	Median (range): 1.5 (0.1–7.1)	CY+ATG	8/40 (20%)	NSD ($p = 0.462$)
			CY+FLUD+ATG	6/43 (14%)	
cGVHD	Extensive in evaluable patients	Median (range): 1.5 (0.1–7.1)	CY+ATG	4/30 (13%)	NSD ($p = 0.692$)
			CY+FLUD+ATG	3/37 (8%)	
cGVHD	Overall in evaluable patients	Median (range): 1.5 (0.1–7.1)	CY+ATG	5/30 (17%)	NSD ($p = 0.961$)
			CY+FLUD+ATG	6/37 (16%)	
Veno-occlusive disease	Sinusoidal obstructive syndrome	Median (range): 1.5 (0.1–7.1)	CY+ATG	5/40 (13%)	NSD ($p = 0.101$)
			CY+FLUD+ATG	1/43 (2%)	
Pneumonia or pneumonitis	—	Median (range): 1.5 (0.1–7.1)	CY+ATG	6/40 (15%)	NR
			CY+FLUD+ATG	2/43 (5%)	
CMV	—	Median (range):	CY+ATG	24/40 (60%)	NR

Summary of results for all reported outcomes: Kim et al. (2012)[42, 43]					
Outcome	**Outcome definition**	**Follow-up time**	**Treatment regimen**	**Findings by treatment arm**	**Study conclusions**
antigenemia		1.5 (0.1–7.1)	CY+FLUD+ATG	23/43 (54%)	
Any regimen-related toxicity[a]	—	Median (range): 1.5 (0.1–7.1)	CY+ATG	22/40 (55%)	**Significant difference (p = 0.003)**
			CY+FLUD+ATG	10/43 (23%)	
Pulmonary complications[b]	—	Median (range): 1.5 (0.1–7.1)	CY+ATG	14/40 (35%)	**Significant difference (p = 0.011)**
			CY+FLUD+ATG	5/43 (12%)	

[a]included sinusoidal obstructive syndrome, hematuria, pulmonary complications, and death from any cause
[b]included bacterial pneumonia, pulmonary TB reactivation, pneumonitis caused by aGVHD, pulmonary edema caused by acute renal failure, and other unknown causes of pulmonary complications
ATG = anti-thymocyte globulin; CY = cyclophosphamide; FLUD = fludarabine

Appendix 5: Summary of Results for Subgroups Reported for All Outcomes

Summary of results for subgroups presented in studies evaluating treatment of GVHD						
Author (Year)	**Patient sub-group**	**Post-transplant follow-up (years)**	**Outcome definition**	**Comparison (sample size)**	**Group risk n (%)**	**OR (95% CI) or study conclusions**
Overall mortality						
Kim (2012)[42, 43]	Matched sibling donors	Median (range): 1.5 (0.1–7.1)	There were no relapses, so overall mortality was equal to non-relapse mortality	CY+ATG (26) vs. CY+ATG+FLUD (27)*	4 (15%) vs. 3 (11%)	1.45 (0.29–7.24)
Kim (2012)[42, 43]	Alternative donors	Median (range): 1.5 (0.1–7.1)	There were no relapses, so overall mortality was equal to non-relapse mortality	CY+ATG (14) vs. CY+ATG+FLUD (16)*	4 (29%) vs. 3 (19%)	1.73 (0.31–9.57)
Non-relapse mortality						
Kim (2012)[42, 43]	Matched sibling donors	Median (range): 1.5 (0.1–7.1)	There were no relapses, so overall mortality was equal to non-relapse mortality	CY+ATG (26) vs. CY+ATG+FLUD (27)*	4 (15%) vs. 3 (11%)	1.45 (0.29–7.24)
Kim (2012)[42, 43]	Alternative donors	Median (range): 1.5 (0.1–7.1)	There were no relapses, so overall mortality was equal to non-relapse mortality	CY+ATG (14) vs. CY+ATG+FLUD (16)*	4 (29%) vs. 3 (19%)	1.73 (0.31–9.57)
Blume (1993)[32]	AML	Median (range): 2.6 (0.8–4.8)	—	ETP+TBI (22) vs. BU+CY (18)*	10 (46%) vs. 7 (39%)	1.31 (0.37–4.64)
Blume (1993)[32]	ALL	Median (range): 2.6 (0.8–4.8)	—	ETP+TBI (25)* vs. BU+CY (23)	4 (16%) vs. 6 (26%)	1.85 (0.45–7.65)
Blume (1993)[32]	CML	Median (range): 2.6 (0.8–4.8)	—	ETP+TBI (14)* vs. BU+CY (20)	3 (21%) vs. 8 (40%)	2.44 (0.51–11.62)
Blume (1993)[32]	Good risk	Median (range): 2.6 (0.8–4.8)	—	ETP+TBI (20)* vs. BU+CY (27)	6 (30%) vs. 10 (37%)	1.37 (0.40–4.72)
Blume (1993)[32]	Poor risk	Median (range): 2.6 (0.8–4.8)	—	ETP+TBI (35)* vs. BU+CY (32)	11 (31%) vs. 11 (34%)	1.14 (0.41–3.17)
Bornhauser (2011)[33]	>40 years of age	1-year cumulative incidence	—	CY+TBI (?) vs. FLUD+TBI (?)	20% vs. 5%	**p = 0.01**
Ringdén (1999)[49-51]	High risk	7-year Kaplan-	—	BU+CY (29) vs.	64% vs. 22%	**p = 0.004**

Summary of results for subgroups presented in studies evaluating treatment of GVHD						
Author (Year)	**Patient sub-group**	**Post-transplant follow-up (years)**	**Outcome definition**	**Comparison (sample size)**	**Group risk n (%)**	**OR (95% CI) or study conclusions**
		Meier		CY+TBI (18)		
Ringdén (1999)[49-51]	Advanced disease (beyond first remission or first chronic phase)	3-year Kaplan-Meier	—	BU+CY (29) vs. CY+TBI (18)	62% vs. 12%	**p < 0.002**
Extensive chronic GVHD in patients alive at 100 days post-transplant						
Kim (2012)[42, 43]	Matched-sibling donors	Median (range): 1.5 (0.1–7.1)	Extensive cGVHD, according to Seattle criteria	CY+ATG (21) vs. CY+ATG+FLUD (24)*	2 (10%) vs. 1 (4%)	2.42 (0.20–28.80)
Kim (2012)[42, 43]	Alternative donors	Median (range): 1.5 (0.1–7.1)	Extensive cGVHD, according to Seattle criteria	CY+ATG (9) vs. CY+ATG+FLUD (13)*	2 (22%) vs. 1 (8%)	3.43 (0.26–45.03)
Relapse						
Blume (1993)[32]	AML	Median (range): 2.6 (0.8–4.8)	—	ETP+TBI (22) vs. BU+CY (18)*	8 (36%) vs. 6 (33%)	1.14 (0.31–4.23)
Blume (1993)[32]	ALL	Median (range): 2.6 (0.8–4.8)	—	ETP+TBI (25)* vs. BU+CY (23)	12 (48%) vs. 12 (52%)	1.18 (0.38–3.67)
Blume (1993)[32]	CML	Median (range): 2.6 (0.8–4.8)	—	ETP+TBI (14) vs. BU+CY (20)*	4 (29%) vs. 3 (15%)	2.27 (0.42–12.27)
Blume (1993)[32]	Good risk	Median (range): 2.6 (0.8–4.8)	—	ETP+TBI (20)* vs. BU+CY (27)	3 (15%) vs. 8 (30%)	2.39 (0.54–10.48)
Blume (1993)[32]	Poor risk	Median (range): 2.6 (0.8–4.8)	—	ETP+TBI (35) vs. BU+CY (32)*	21 (60%) vs. 13 (41%)	2.19 (0.83–5.83)
Liu (2013)[47]	AML	Unclear (possibly 5 years) cumulative incidence	Sum of hematologic and genetic relapse	BU+CY (?) vs. BU+FLUD (?)	? vs. ?	**RR = 2.22 (1.15–4.29) p = 0.018 (reported by authors)**
Ringdén (1999)[49-51]	Early disease	3-year Kaplan-Meier	—	BU+CY (59) vs. CY+TBI (61)	14% vs. 25%	p = 0.29
Ringdén (1999)[49-51]	Advanced disease (beyond first remission or first chronic phase)	3-year Kaplan-Meier	—	BU+CY (29) vs. CY+TBI (18)	47% vs. 31%	p = 0.41
Ringdén (1999)[49-51]	Early disease	7-year Kaplan-Meier	—	BU+CY (59) vs. CY+TBI (61)	21% vs. 27%	NSD

Summary of results for subgroups presented in studies evaluating treatment of GVHD						
Author (Year)	**Patient sub-group**	**Post-transplant follow-up (years)**	**Outcome definition**	**Comparison (sample size)**	**Group risk n (%)**	**OR (95% CI) or study conclusions**
Ringdén (1999)[49-51]	Advanced disease (beyond first remission or first chronic phase)	7-year Kaplan-Meier	—	BU+CY (29) vs. CY+TBI (18)	50% vs. 36%	NSD
Ringdén (1999)[49-51]	Adults only	3-year cumulative incidence	—	BU+CY (73) vs. CY+TBI (67)	23% vs. 24%	NSD
Ringdén (1999)[49-51]	Children only	3-year cumulative incidence	—	BU+CY (15) vs. CY+TBI (12)	16% vs. 33%	NSD
Veno-occlusive disease						
Kim (2012)[42, 43]	Matched-sibling donors	Median (range): 1.5 (0.1–7.1)	—	CY+ATG (26) vs. CY+ATG+FLUD (27)	3 (12%) vs. 0 (0%)	Not estimable
Kim (2012)[42, 43]	Alternative donors	Median (range): 1.5 (0.1–7.1)	—	CY+ATG (14) vs. CY+ATG+FLUD (16)*	2 (14%) vs. 1 (6%)	2.50 (0.20–31.00)
Infection—Pneumonia/pneumonitis						
Kim (2012)[42, 43]	Matched-sibling donors	Median (range): 1.5 (0.1–7.1)	—	CY+ATG (26) vs. CY+ATG+FLUD (27)*	4 (15%) vs. 2 (7%)	2.27 (0.38–13.63)
Kim (2012)[42, 43]	Alternative donors	Median (range): 1.5 (0.1–7.1)	—	CY+ATG (14) vs. CY+ATG+FLUD (16)*	2 (14%) vs. 0 (0%)	Not estimable
Infection—CMV antigenemia						
Kim (2012)[42, 43]	Matched-sibling donors	Median (range): 1.5 (0.1–7.1)	—	CY+ATG (26) vs. CY+ATG+FLUD (27)*	16 (62%) vs. 15 (56%)	1.28 (0.43–3.83)
Kim (2012)[42, 43]	Alternative donors	Median (range): 1.5 (0.1–7.1)	—	CY+ATG (14) vs. CY+ATG+FLUD (16)*	8 (57%) vs. 8 (50%)	1.33 (0.32–5.64)
Any regimen-related toxicity[a]						
Kim (2012)[42, 43]	Matched-sibling donors	Median (range): 1.5 (0.1–7.1)	—	CY+ATG (26) vs. CY+ATG+FLUD (27)*	16 (62%) vs. 6 (22%)	**5.60 (1.68–18.65)**
Kim (2012)[42, 43]	Alternative donors	Median (range): 1.5 (0.1–7.1)	—	CY+ATG (14) vs. CY+ATG+FLUD (16)*	6 (43%) vs. 4 (25%)	2.25 (0.48–10.60)

Summary of results for subgroups presented in studies evaluating treatment of GVHD						
Author (Year)	**Patient sub-group**	**Post-transplant follow-up (years)**	**Outcome definition**	**Comparison (sample size)**	**Group risk n (%)**	**OR (95% CI) or study conclusions**
Pulmonary complications[b]						
Kim (2012)[42, 43]	Matched-sibling donors	Median (range): 1.5 (0.1–7.1)	—	CY+ATG (26) vs. CY+ATG+FLUD (27)*	9 (35%) vs. 2 (7%)	**6.62 (1.27–34.51)**
Kim (2012)[42, 43]	Alternative donors	Median (range): 1.5 (0.1–7.1)	—	CY+ATG (14) vs. CY+ATG+FLUD (16)*	5 (36%) vs. 2 (13%)	3.89 (0.62–24.52)

[a]included sinusoidal obstructive syndrome, hematuria, pulmonary complications, and death from any cause
[b]included bacterial pneumonia, pulmonary TB reactivation, pneumonitis caused by aGVHD, pulmonary edema caused by acute renal failure, and other unknown causes of pulmonary complications
*denotes reference group for odds ratio calculation

ATG = anti-thymocyte globulin; BU = busulfan; CY = cyclophosphamide; ETP = etoposide; FLUD = fludarabine; TBI = total body irradiation

Appendix 6: Summary of Findings from Random Effects NMAs for the Review

100-day Mortality

BU+FLUD				
0.66 (0.05 – 5.53)	**ETP+TBI**			
0.50 (0.00 – 709.00)	0.75 (0.00 – 1283.86)	**MELPH+TBI**		
0.46 (0.05 – 2.70)	0.70 (0.11 – 4.68)	0.93 (0.00 – 471.60)	**CY+TBI**	
0.36 (0.05 – 1.55)	0.54 (0.10 – 2.96)	0.72 (0.00 – 375.10)	0.78 (0.31 – 1.84)	**BU+CY**

1-Year Mortality

CY+TBI				
0.84 (0.13 – 4.95)	**MELPH+TBI**			
0.67 (0.19 – 2.55)	0.81 (0.09 – 8.01)	**BU+FLUD**		
0.61 (0.26 – 1.44)	0.73 (0.10 – 5.62)	0.91 (0.33 – 2.38)	**BU+CY**	
0.43 (0.07 – 2.79)	0.52 (0.04 – 6.75)	0.64 (0.09 – 4.43)	0.71 (0.13 – 3.76)	**ETP+TBI**

2-Year Mortality

CY+TBI				
0.99 (0.11 – 9.38)	MELPH+TBI			
0.50 (0.05 – 5.43)	0.51 (0.02 – 12.43)	ETP+TBI		
0.51 (0.17 – 1.49)	0.52 (0.04 – 5.85)	1.01 (0.12 – 8.53)	BU+CY	
0.37 (0.06 – 2.65)	0.38 (0.02 – 7.13)	0.74 (0.06 – 10.82)	0.73 (0.17 – 3.62)	BU+FLUD

4-Year Mortality

ETP+TBI				
0.61 (0.07 – 4.86)	CY+TBI			
0.59 (0.04 – 9.61)	0.99 (0.16 – 5.86)	MELPH+TBI		
0.37 (0.06 – 2.42)	0.62 (0.26 – 1.44)	0.62 (0.09 – 4.47)	BU+CY	
0.32 (0.03 – 3.17)	0.53 (0.13 – 2.41)	0.54 (0.06 – 5.68)	0.86 (0.27 – 3.08)	BU+FLUD

5-Year Mortality

ETP+TBI				
0.82 (0.05 – 13.38)	MELPH+TBI			
0.58 (0.07 – 4.87)	0.71 (0.12 – 4.23)	CY+TBI		
0.46 (0.05 – 4.55)	0.57 (0.06 – 6.05)	0.79 (0.19 – 3.90)	BU+FLUD	
0.37 (0.05 – 2.42)	0.45 (0.06 – 3.43)	0.64 (0.24 – 1.69)	0.81 (0.24 – 2.39)	BU+CY

Findings, Veno-Occlusive Disease

CY+TBI			
0.53 (0.04 – 6.80)	MELPH+TBI		
0.40 (0.01 – 12.57)	0.75 (0.01 – 55.81)	BU+FLUD	
0.15 (0.02 – 1.03)	0.27 (0.01 – 6.92)	0.37 (0.02 – 5.14)	BU+CY

Findings, Acute GVHD

TBI								
0.77 (0.08 – 8.84)	ATG+TLI							
0.59 (0.04 – 8.26)	0.76 (0.04 – 12.15)	BU+FLUD						
0.53 (0.12 – 2.40)	0.70 (0.10 – 3.96)	0.90 (0.10 – 8.08)	FLUD+TBI					
0.49 (0.03 – 6.96)	0.63 (0.04 – 9.83)	0.83 (0.10 – 7.04)	0.93 (0.10 – 8.25)	ETP+TBI				
0.39 (0.04 – 3.36)	0.50 (0.04 – 5.08)	0.66 (0.15 – 3.08)	0.73 (0.15 – 3.59)	0.78 (0.16 – 3.62)	BU+CY			
0.32 (0.04 – 2.46)	0.40 (0.04 – 3.65)	0.54 (0.10 – 2.95)	0.60 (0.15 – 2.41)	0.64 (0.12 – 3.52)	0.82 (0.40 – 1.67)	CY+TBI		
0.22 (0.03 – 1.75)	0.30 (0.03 – 2.65)	0.39 (0.03 – 5.03)	0.43 (0.10 – 1.71)	0.46 (0.03 – 6.22)	0.59 (0.07 – 4.74)	0.72 (0.10 – 5.15)	ATG+BU+FLUD	
0.12 (0.01 – 1.68)	0.16 (0.01 – 2.36)	0.21 (0.02 – 2.07)	0.23 (0.03 – 1.91)	0.25 (0.02 – 2.53)	0.32 (0.06 – 1.73)	0.39 (0.08 – 1.87)	0.55 (0.04 – 6.66)	MELPH+TBI

Findings, Relapse at 2-3 Years

BU+CY			
0.70 (0.08 – 6.47)	ETP+TBI		
0.71 (0.21 – 2.07)	1.01 (0.08 – 10.93)	CY+TBI	
0.68 (0.15 – 3.20)	0.97 (0.07 – 14.23)	0.97 (0.15 – 6.98)	BU+FLUD

Findings, Relapse at 4-5 Years

BU+FLUD+ATG		
0.32 (0.02 – 4.20)	FLUD+TBI	
0.08 (0.00 – 3.29)	0.27 (0.02 – 3.75)	ATG+TLI

Appendix 7: PRISMA NMA checklist

PRISMA NMA Checklist of Items to Include When Reporting a Systematic Review Involving a Network Meta-analysis

Section/Topic	Item #	Checklist Item	Reported on Page #
TITLE			
Title	1	Identify the report as a systematic review *incorporating a network meta-analysis (or related form of meta-analysis).*	1
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: **Background:** main objectives **Methods:** data sources; study eligibility criteria, participants, and interventions; study appraisal; and *synthesis methods, such as network meta-analysis.* **Results:** number of studies and participants identified; summary estimates with corresponding confidence/credible intervals; *treatment rankings may also be discussed. Authors may choose to summarize pairwise comparisons against a chosen treatment included in their analyses for brevity.* **Discussion/Conclusions:** limitations; conclusions and implications of findings. **Other:** primary source of funding; systematic review registration number with registry name.	2
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known, *including mention of why a network meta-analysis has been conducted.*	11
Objectives	4	Provide an explicit statement of questions being addressed, with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	12

METHODS			
Protocol and registration	5	Indicate whether a review protocol exists and if and where it can be accessed (e.g., Web address); and, if available, provide registration information, including registration number.	12
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale. *Clearly describe eligible treatments included in the treatment network, and note whether any have been clustered or merged into the same node (with justification).*	12
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	13
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	Appendix 1
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	14
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	14
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	14
Geometry of the network	**S1**	Describe methods used to explore the geometry of the treatment network under study and potential biases related to it. This should include how the evidence base has been graphically summarized for presentation, and what characteristics were compiled and used to describe the evidence base to readers.	14
Risk of bias within individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	23, 25

Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means). *Also describe the use of additional summary measures assessed, such as treatment rankings and surface under the cumulative ranking curve (SUCRA) values, as well as modified approaches used to present summary findings from meta-analyses.*	16
Planned methods of analysis	14	Describe the methods of handling data and combining results of studies for each network meta-analysis. This should include, but not be limited to: • *Handling of multi-arm trials;* • *Selection of variance structure;* • *Selection of prior distributions in Bayesian analyses; and* • *Assessment of model fit.*	16-17
Assessment of Inconsistency	**S2**	Describe the statistical methods used to evaluate the agreement of direct and indirect evidence in the treatment network(s) studied. Describe efforts taken to address its presence when found.	18
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	NA
Additional analyses	16	Describe methods of additional analyses if done, indicating which were pre-specified. This may include, but not be limited to, the following: • Sensitivity or subgroup analyses; • Meta-regression analyses; • *Alternative formulations of the treatment network; and* • *Use of alternative prior distributions for Bayesian analyses (if applicable).*	18
RESULTS†			

Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	19
Presentation of network structure	**S3**	Provide a network graph of the included studies to enable visualization of the geometry of the treatment network.	Throughout results
Summary of network geometry	**S4**	Provide a brief overview of characteristics of the treatment network. This may include commentary on the abundance of trials and randomized patients for the different interventions and pairwise comparisons in the network, gaps of evidence in the treatment network, and potential biases reflected by the network structure.	24
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	21-22
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment.	23, 25
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: 1) simple summary data for each intervention group, and 2) effect estimates and confidence intervals. *Modified approaches may be needed to deal with information from larger networks.*	Upon request
Synthesis of results	21	Present results of each meta-analysis done, including confidence/credible intervals. *In larger networks, authors may focus on comparisons versus a particular comparator (e.g. placebo or standard care), with full findings presented in an appendix. League tables and forest plots may be considered to summarize pairwise comparisons.* If additional summary measures were explored (such as treatment rankings), these should also be presented.	26-54
Exploration for inconsistency	**S5**	Describe results from investigations of inconsistency. This may include such information as measures of model fit to compare consistency and inconsistency models, *P* values from statistical tests, or summary of inconsistency estimates from different parts of the treatment network.	NA (no closed loops)
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies for the evidence base being studied.	NA
Results of additional analyses	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression analyses, *alternative network geometries studied, alternative choice of prior distributions for Bayesian analyses,* and so forth).	54

DISCUSSION			
Summary of evidence	24	Summarize the main findings, including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy-makers).	54
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review level (e.g., incomplete retrieval of identified research, reporting bias). *Comment on the validity of the assumptions, such as transitivity and consistency. Comment on any concerns regarding network geometry (e.g., avoidance of certain comparisons).*	55
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	55
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review. This should also include information regarding whether funding has been received from manufacturers of treatments in the network and/or whether some of the authors are content experts with professional conflicts of interest that could affect use of treatments in the network.	55

PICOS = population, intervention, comparators, outcomes, study design.

* Text in italics indicates wording specific to reporting of network meta-analyses that has been added to guidance from the PRISMA statement.

† Authors may wish to plan for use of appendices to present all relevant information in full detail for items in this section.