Health Quality Ontario

Ontario Health Technology Assessment Series

Minimal Residual Disease Evaluation in Childhood Acute Lymphoblastic Leukemia: A Clinical Evidence Review

KEY MESSAGES

Leukemia is a blood cancer that is the most common childhood cancer in Canada. The most common type of leukemia is called acute lymphoblastic leukemia. Treatment is now very successful at curing patients because doctors can forecast a child's chance of getting better or having leukemia return (relapse) based on their clinical or physical characteristics. By determining if a child has a high or low risk of relapse, doctors can choose the best treatment for each child. But current methods to determine a child's risk of relapse are not perfect, and about 20% of children who are expected to have good outcomes (standard risk) still relapse.

Relapse might be caused by cancer cells left over in the bone marrow at extremely low levels after treatment, even though the disease seems to be in remission. This extremely low level of cancer cells is called minimal residual disease. Selecting treatment for standard-risk patients according to how many leukemia cells are left once they reach remission can improve how well patients recover. Previous studies have found minimal residual disease is related to event-free survival, where relapses are a common event. This study reviewed the evidence that minimal residual disease can provide information about relapse in acute lymphoblastic leukemia, and be used as a way to choose treatment.

Although researchers study minimal residual disease in different ways, they find that minimal residual disease is a valuable test in acute lymphoblastic leukemia. If levels of minimal residual disease are high at the end of the first or second phases of treatment in newly diagnosed patients, at the end of the first phase of treatment for relapsed patients, and before a stem cell transplant, patients are more likely to relapse. Researchers also found that standard-risk patients with low levels of minimal residual disease can still get better with reduced treatment. Similarly, standard-risk patients with high levels of minimal residual disease get better when they are given more intense treatment than usual.

March 8, 2016
VOL. 16, NO. 7

Ontario
Health Quality Ontario

HEALTH TECHNOLOGY ASSESSMENT AT HEALTH QUALITY ONTARIO

ABSTRACT

Background

Leukemia accounts for nearly a third of childhood cancers in Canada, with acute lymphoblastic leukemia (ALL) comprising nearly 80% of cases. Identification of prognostic factors that allow risk stratification and tailored treatment have improved overall survival. However, nearly a quarter of patients considered standard risk on the basis of conventional prognostic factors still relapse, and relapse is associated with increased morbidity and mortality. Relapse is thought to result from extremely low levels of leukemic cells left over once complete remission is reached, termed minimal residual disease (MRD). Poor event-free survival (EFS) as well as overall survival for those who are classified as MRD-positive have been substantiated in seminal studies demonstrating the prognostic value of MRD for EFS in the past few decades. This review sought to further elucidate the relationship between MRD and EFS by looking at relapse, the primary determinant of EFS and the biological mechanism through which MRD is thought to act. This evidence review aimed to ascertain whether MRD is an independent prognostic factor for relapse and to assess the effect of MRD-directed treatment on patient-important outcomes in childhood ALL.

Methods

Large prospective cohort studies with a priori multivariable analysis that includes potential confounders are required to draw confirmatory conclusions about the independence of a prognostic factor. Data on the prognostic value of MRD for relapse measured by molecular methods (polymerase chain reaction [PCR] of immunoglobulin or T-cell receptor rearrangements) or flow cytometry for leukemia-associated immunophenotypes or difference-from-normal approach were abstracted from included studies. Relevant data on relapse, EFS, and overall survival were abstracted from randomized controlled trials (RCTs) evaluating the effect of MRD-directed treatment.

Results

A total of 2,832 citations were reviewed, of which 12 studies were included in this review. All cohort studies evaluating MRD as a prognostic factor for relapse found significant independent value when added to various existing prognostic factors. Seven studies showed prognostic value of MRD measured at the end of induction therapy and two at the end of consolidation therapy in de novo ALL, one study in relapsed ALL after re-induction therapy, and three studies before hematopoietic stem cell transplant. One large RCT in standard-risk patients found no compromise to outcomes when reducing treatment in MRD-negative patients, and also showed a 45% reduction in relapse risk and nearly 40% benefit in EFS when escalating treatment in MRD-positive patients.

Conclusions

Minimal residual disease is an independent prognostic factor for relapse in childhood ALL. Relapse is a key determinant of EFS and patients' quality of life. Treatment selected on the basis of MRD status appears to improve outcomes.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

BACKGROUND

Objective of Analysis

This evidence review aimed to (1) ascertain whether minimal residual disease (MRD) evaluation is an independent prognostic factor for relapse in children with acute lymphoblastic leukemia (ALL), and (2) assess the effect of MRD-directed treatment on patient-important outcomes for children with ALL.

Clinical Need and Target Population

Description of Disease

Leukemia is a relatively rare form of cancer that affects the cells that form the blood.[1] Although leukemia comprises a small proportion of adult cancers (about 3%), it is the most commonly diagnosed childhood cancer in Canada, accounting for 32% of diagnoses among children 0 to 14 years old.[2] The disease occurs when blood stem cells produced in the bone marrow do not mature properly and are overproduced.[1] There are two main types of acute childhood leukemia, named according to the type of blood stem cell that is affected: ALL and acute myeloid leukemia (AML). Table 1 outlines normal blood cell development from blood stem cells.

Table 1: Normal Blood Cell Development

Type of Blood Stem Cell	Intermediate Cell Type	Mature Cell Types	Normal Mature Cell Function(s)
Lymphoid	Lymphoblast	B lymphocyte	Make antibodies to fight infection
	Lymphoblast	T lymphocyte	Assist B lymphocytes to make antibodies
	Lymphoblast	Natural killer cell	Attack cancer cells and viruses
Myeloid	Myeloblast	Granulocytes • Eosinophil • Neutrophil • Basophil	Fight infection and disease
		Platelet	Form blood clots to stop bleeding
		Red blood cell	Carry oxygen and other nutrients to all tissues of the body

Source: National Cancer Institute.[3]

In ALL, lymphoid stem cells develop into lymphoblasts or into poorly functioning B lymphocytes or T lymphocytes instead of mature white blood cells that provide immune system support.[3] Acute lymphoblastic leukemia is further categorized by cell lineage (or immunophenotype) as T-cell ALL (approximately 15% of cases) or B-cell ALL (up to 85% of cases) based on whether the leukemic cells originated as T- or B-lymphocytes, respectively.[3] Subtypes of B-cell ALL are defined by the maturity of the lymphoblasts, but the vast majority of ALL comes from precursor-B lymphocytes and thus is referred to as B-cell ALL throughout this document for simplicity. In contrast, when myeloid stem cells develop into myeloblasts instead of red blood cells, white blood cells, or platelets, they cause AML.[4] The accumulation of leukemic cells in the blood and bone marrow reduces room for healthy blood cells and can lead to leukemic cells spreading to other areas of the body.[4] Possible symptoms of leukemia include bleeding, bruising easily, fatigue, frequent infections, fever, anemia, and other flu-like symptoms.[3]

Diagnosis takes place once symptoms appear and can include physical examination, complete blood count, and a biopsy of bone marrow.[3] Bone marrow samples typically are assessed first via morphology (i.e., microscopic examination for signs of cancer) and may subsequently undergo microscopic examination to view chromosomal changes.[3] Chromosomal analysis can include fluorescence in situ hybridization (FISH) analysis to identify specific strings of deoxyribonucleic acid (DNA) on the cells.[4] Analysis of the chromosomes is key to identifying and characterizing the cancer cell line that is present and to distinguish the leukemic cells (or clones).[5]

Treatment for de novo ALL (new cases) broadly consists of a series of phases of therapy: remission induction, consolidation, intensification, and maintenance (Figure 1).[3] The duration and drug combination in each treatment phase can vary between clinical protocols, but most can be simplified into the same overall components. The goal is first, via induction therapy, to eradicate (> 99%) leukemic cells to bring about complete remission (defined as < 5% leukemic cells assessed by light microscopy of a bone marrow sample). Although the exact regimens differ between treatment protocols, some key drugs are invariably administered because they are effective. Consequently, induction rates for complete remission of childhood ALL overall are in the order of 97% or higher.[6] Induction is followed by the phases of consolidation and intensification to kill any remaining leukemic cells to prevent relapse, and finally maintenance, which is a type of continuation therapy that aims to prohibit growth of leukemic cells that could regrow and cause relapse.[3] Patients who do not achieve complete remission (who have induction failure) or relapse may have the option for hematopoietic stem cell transplant (HSCT) following re-induction therapy (Figure 2).

Figure 1: Conceptual Model of Role of Minimal Residual Disease (MRD) in Acute Lymphoblastic Leukemia Patient Outcome
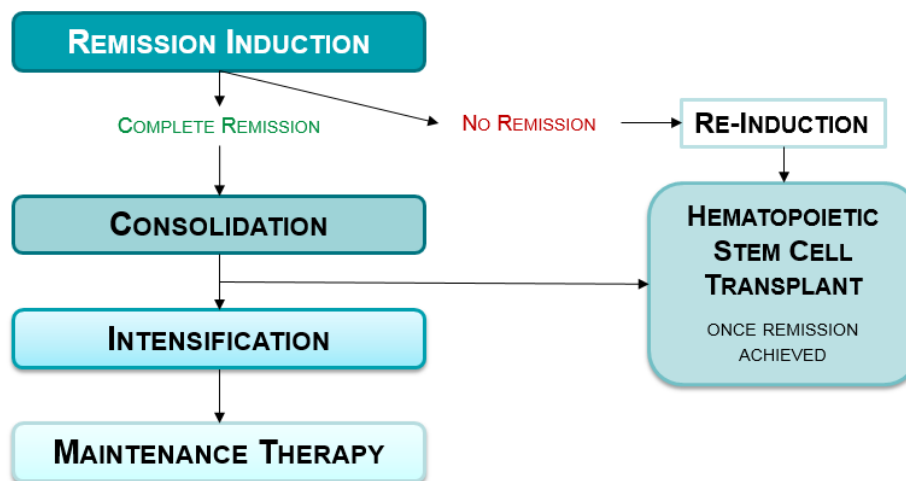
Figure 2: Simplified Schema of Acute Lymphoblastic Leukemia Treatment Phases

Relapse prevention is of paramount importance because relapse after remission induction therapy is responsible for both substantial morbidity and nearly 20% of deaths from leukemia.[7,8] Relapse is determined by morphology (light microscope) and is the return of leukemia after complete remission, most commonly in the bone marrow with or without involvement of another body site.[9] Despite treatment advances since the 1950s and 1960s, overall approximately 25% of patients with ALL relapse.[10-12] Extramedullary (outside the bone marrow) or combined relapse can include the central nervous system (CNS) or testis, in part because chemotherapy tends to poorly penetrate these areas (referred to as sanctuary sites).[9] Central nervous system relapse can result from leukemic cells from the bone marrow of the skull entering into the cerebrospinal fluid, nerve roots, or veins then into the subarachnoid space (of the brain or spinal cord) and is seen in approximately 3% to 8% of patients with ALL, mostly in clinically high-risk patients.[13] Testicular relapse occurs in 1% to 2% of boys[3] and is more likely to be isolated (without bone marrow involvement), as the testes are an anatomical site where leukemic cells are known to remain.[9]

Relapse can be classified as early (< 36 months from diagnosis) or late (> 36 months from diagnosis), as well as by timing (on-therapy or off-therapy) and site (bone marrow, extramedullary, or combined).[9] Patients with early bone marrow relapse achieve complete remission a second time less frequently than patients with late bone marrow relapse (68% vs. 96%).[14] Bone marrow relapse within 5 years of diagnosis is broadly accepted as resulting from residual leukemic blasts (MRD) that persist despite reaching remission through therapy (Figure 1).[9,15] In some cases, the leukemic cells (leukemic clones) identified at diagnosis are the same types as at relapse, whereas in other cases a new type of leukemic cell population is found at relapse (referred to as clonal evolution).[15] It is not entirely clear whether the cells that persist have acquired chemoresistance owing to mutations resulting from exposure to therapy or whether they have an innate treatment resistance.[9,15] A seminal study found that at least some of the leukemic clones present at relapse were present at diagnosis in nearly all cases.[16]

Outcomes for relapsed ALL are poor; survival varies from 10% to 60% for relapsed patients with high- and standard-risk characteristics, respectively.[6] Patients who experience bone marrow relapse within the 12 months of complete remission and who then undergo chemotherapy or HSCT have been observed to have overall survival and event-free survival (EFS) at 5 years as low as 18% and 16%, respectively.[17] In contrast, among patients with ALL who survive and remain in complete remission for 2 years following HSCT, the 10-year overall survival is far superior, about 84%.[18] The eradication of existing leukemic blasts and prevention of relapse, to enable the restoration of normal hematopoiesis, are the core goals of therapy.

To this end, treatment could include chemotherapy and radiation therapy, as well as potentially targeted therapy (i.e., treatments that attack cancer cells and preserve healthy cells).[3] To deliver (intrathecal) chemotherapy to eradicate any leukemic cells in the brain or spinal cord, CNS prophylaxis might also be administered.[4] Various groups (Table 2) have conducted international clinical trials in pediatric ALL to identify the optimal combination, duration, dosage, and timing of chemotherapy and, for high-risk or relapsed patients, of HSCT.

Table 2: Examples of Research Groups Testing Acute Lymphoblastic Leukemia Treatment Protocols

Abbreviation	Full Name	Countries
AIEOP	Associazione Italiana Ematologia Oncologia Pediatrica	Italy
BFM	Berlin–Frankfurt–Münster study group	Germany, Austria, Switzerland
COG	Children's Oncology Group[a]	Australia, Canada, Ireland, Mexico, New Zealand, Saudi Arabia, Switzerland, United States
DCOG	Dutch Childhood Oncology Group	The Netherlands
DFCI	Dana-Faber Cancer Institute consortium	United States
EORTC-CLG	European Organization for Research and Treatment of Cancer–Children's Leukemia Group	France, Belgium, Portugal
NOPHO	NOPHO, Nordic Society of Pediatric Hematology and Oncology	Denmark, Finland, Iceland, Norway, Sweden
PETHEMA	Programa de Estudio Tratamiento de las Hemopatías Malignas	Spain

[a]Formed in 2000 via merger of four predecessor organizations: Children's Cancer Group (CCG), the Pediatric Oncology Group (POG), the Intergroup Rhabdomyosarcoma Study Group (IRSG), and the National Wilms' Tumor Study Group (NWTS).
Source: Kotecha et al, 2014.[19]

High-risk patients (those with relapsed ALL or who do not reach complete remission [induction failure]) undergo re-induction therapy to bring about a second complete remission, which is successful in about 85% of cases overall.[3] Treatment regimens (referred to as salvage or rescue therapy) for high-risk patients tend to include drugs similar to those in first-line therapy, albeit in different doses and frequencies.[9] For instance, the Children's Oncology Group has developed a suitable re-induction platform for relapsed B-cell ALL that will enable refinement of relapse therapy.[14] How a patient responds to each phase helps inform the subsequent treatment plan.

Historically, the potential for allogenic (from a non–identical-twin donor) HSCT was based on the availability of a sibling donor[20]; however, HSCT is now commonly performed with matched, unrelated donors.[6] A high-risk procedure, HSCT requires the complete eradication (conditioning) of immature blood cells by high-intensity chemotherapy or irradiation, which leaves patients susceptible to infection or bleeding. Long-term implications (such as graft-versus-host disease [GVHD], where donated white blood cells can attack the patient's body) are a serious and potentially long-term risk.[21] Most deaths after HSCT occur within the first 2 years because of relapse, GVHD, infection, or other HSCT toxicities.[18] Acute or chronic GVHD occurs in 30% to 60% of new HSCT recipients, with nearly 50% mortality.[22]

A system of risk stratification to identify the likelihood of poor outcomes and necessary treatment course is a cornerstone of ALL treatment. Pediatric patients with ALL are classified as standard- or high-risk on the basis of age and white blood cell count (together known as the National Cancer Institute [NCI] criteria), though other factors also affect the risk group: such as sex, CNS involvement, early response to treatment, and cytogenetics (Table 3).[3] These are established prognostic factors for outcomes in ALL that have been accepted into standard practice. Acute lymphoblastic leukemia treatment, including risk stratification, has evolved from the mid-1900s to now achieve excellent outcomes with 5-year overall survival on the order of 90%.[19]

Table 3: Factors Informing Patient Risk Stratification in Pediatric Patients with De Novo Acute Lymphoblastic Leukemia

Risk Group	Age, Years	White Blood Cell Count at Diagnosis	Other Influential Factors
Standard[a] risk	1 to < 10	< 50 000/μL	• T-cell or B-cell origin (immunophenotype) • Chromosomal changes in lymphocyte genes (cytogenetics)
High risk	≥ 10	≥ 50 000/μL	• Rapidity and magnitude of drop in leukemic cell count after induction therapy • Presence of leukemic cells in cerebrospinal fluid

[a]May also be referred to as low risk.
Source: National Cancer Institute.[3]

In addition to clinical features, cytogenetics evaluates the presence of chromosomal abnormalities (i.e., number of chromosomes, translocations, amplifications, deletions, or mutations) that are associated with favourable or unfavourable outcomes.[23] Some genetic subpopulations are known to respond differently to treatment (e.g., patients with Down syndrome are at higher risk of treatment-related mortality). Others are at very high risk with extraordinarily poor prognosis, including infants (i.e., < 1 year) and patients with *MLL* gene translocations such as t(4;11) and t(9;22; Philadelphia chromosome–positive [Ph+]) or deletion of the *IKZF1* gene.[23] Adolescents and young adults with ALL are also high risk, with generally poorer outcomes than children.[24] Patients with ALL who are classified as high risk generally receive more intense treatments (e.g., higher doses or more anticancer drugs and potentially HSCT) in order to induce remission and prevent subsequent relapse.[3]

Monitoring the quantity of leukemic cells over the course of treatment is informative about the response to therapy and for choosing subsequent treatment. Early treatment response is also informative and can include bone marrow morphology in the first week or two after the start of induction (favourable if < 5% of cells) or peripheral blood morphology following corticosteroid pretreatment (favourable if < 1,000/μL after 7 days).[3] However, morphological assessment of bone marrow cannot detect concentrations below 1% to 5% of leukemic cells (i.e., 1 in 100 to 1 in 20), nor predict which patients who achieve complete remission will relapse.[20] The presence of leukemic cells that are undetectable by morphology,[25] a measure that has considerable value in identifying preclinical relapse, poor response to treatment, and poorer prognosis, is referred to as minimal residual disease (MRD).

Technology/Technique

Minimal residual disease is when leukemic cells in the bone marrow present at very, very low levels even though the patient is in complete remission (i.e., < 5% leukemic cells in a bone marrow sample). Minimal residual disease is assessed via laboratory testing and is dependent upon the test's ability to recognize leukemic cells. Samples of peripheral blood or bone marrow can be analyzed with either polymerase chain reaction (PCR) or flow cytometry.

In Europe the molecular laboratory evaluation technique PCR is most widely used; it detects gene abnormalities, most commonly antigen-receptor gene rearrangements on immunoglobulin heavy chain (Ig) or T cells (TCR), which distinguish leukemic blasts from healthy cells.[26] For the 90% of patients that can be monitored by Ig/TCR rearrangements (i.e., a target on the leukemic blasts can be identified for the laboratory test), PCR is laborious, as patient-specific primers

must be used because every patient has unique DNA on the regions of interest. Developing patient-specific primers for PCR can take several weeks, after which results can be obtained with relative rapidity.[27] Multiple Ig/TCR rearrangements or targets (e.g., IgH, IgK, TCRγ, TCRδ) are ideal for reproducibility, and the rearrangements can evolve over time, thus "losing" a target for PCR.[26]

Acute lymphoblastic leukemia cells also express certain cell markers that differentiate them from normal cells and as coming from either T- or B-cell precursors.[27] The cell markers occur in a pattern (or profile) unique to each patient and can be identified via flow cytometry in 95% or more of B-cell ALL cases.[27] This pattern is referred to as leukemia-associated immunophenotype and, as with PCR, must be determined at diagnosis to allow for monitoring throughout treatment[26] (for a review, see Campana 2009[27]). A benefit is that the resulting level of MRD can be obtained by flow cytometry within a few hours of collecting the sample to be tested.[27] Laboratory protocols for flow cytometric MRD evaluation are relatively less standardized than for PCR at present, but are in development.[28]

Real-time quantitative PCR for Ig/TCR rearrangements generally has a sensitivity of 1 in 100,000 cells, whereas three- or four-colour flow cytometry for leukemia-associated immunophenotypes usually reaches sensitivity of 1 in 10,000 cells.[29,30] Despite this difference, a trend among studies is to employ a cut-off for MRD positivity (MRD+) of ≥ 0.01% and for negativity (MRD−) of < 0.01%, which aids comparison between studies.[30] Several comparative studies have estimated the concordance between PCR and flow cytometry for MRD detection to be 90% or more of cases, with discrepancies lying at the very lowest level of MRD (where flow cytometry can yield a negative result, while PCR yields a positive result).[30] Experts confirm results of PCR and flow cytometry are equivalent for making clinical decisions. Both PCR and flow cytometric methods are complex and require both specialized expertise and standardization of laboratory techniques for quality assurance.

Previous research has shown that risk stratification helps in tailoring leukemia treatment to patients (i.e., intensified treatment for high-risk and regular treatment for standard-risk patients). Measurement of MRD after induction therapy provides insight into early response to treatment and, because sensitivity is higher, impending relapse can be detected.[27] Some in the field advocate that MRD evaluation provides a more accurate method to redefine leukemia remission than previous assessment methods, detecting much lower leukemic cell concentrations of 0.01% to 1%.[27]

Early reports were among the first to find MRD to be informative about prognosis (e.g., Brisco et al[31]) and were subsequently replicated. In the late 1990s, a previously undetectable difference in overall prognosis in the apparently homogeneous standard-risk group according to MRD results was demonstrated by several research groups.[20] Given that a clinical high-risk patient is categorized as such for many reasons that MRD does not influence (e.g., immunophenotype and cytogenetics), it is in the clinically standard-risk group of patients (~75% of cases[11]) that MRD offers the most value. Various combinations of MRD evaluation at different time points during treatment have been investigated, by such research groups as those in Table 2, for the ability to delineate MRD prognostic subgroups. While the evaluation of MRD has important potential clinical applications, the optimal timing to measure MRD remains unclear. However, general consensus holds that early evaluation is beneficial (e.g., end of induction), although later time points (e.g., end of consolidation) seem to be powerful as well.[12] Thus, MRD might further refine patient-risk stratification and treatment selection, and, ultimately, might improve patient outcomes.

Predictive Markers

Predictive markers identify patients with differential response to treatment (e.g., responders, those likely to experience adverse outcomes) and can be used to guide treatment decisions.[32,33] In oncology, predictive marker studies examine the effect of the treatment on the cancer[34] and can be used for patient selection (up-front predictive marker) or to provide information on early response to therapy (early predictive markers).[35] In contrast, prognostic factors delineate differences between groups of patients in risk of an outcome, given a standard or no treatment (i.e., treatment is held constant).[34,35]

Prognostic Factors

A prognostic factor in oncology is a clinical or biologic measure that predicts clinical course or outcome (e.g., relapse).[32] Simply put, a prognostic factor is a characteristic of the patient or the cancer that affects the patient outcome.[34,35] The potential value of a prognostic factor in ALL is providing information on risk of relapse so that treatment can be modified to minimize chemotherapy toxicity without compromising survival.[35] Though a prognostic factor need not be a predictive marker (identify differential treatment response) to have value, it can be if the association is causal or affects a biological mechanism.[36] To establish and characterize the association between a factor and outcomes, sequential phases of investigation and study designs have been proposed (Table 4).

Table 4: Types of Prognostic Factor Studies

Type of Study	Nature	Goal
Phase I	Exploratory (hypothesis generating)	Identify potential prognostic factors for further investigation
Phase II	Exploratory	Test association between factor and disease outcome
	Confirmatory	Protocol-driven, large prospective studies with a priori hypotheses to examine the strength, direction, and independence of a prognostic factor, accounting for potential confounders

Source: Riley et al, 2009.[32]

Confirmatory studies are considered to be the highest quality and strongest level of evidence in prognostic research.[37,38] In order to establish novel and added prognostic value, studies ought to be designed to include the following features:

- A large, prospective cohort of patients in the same stage of disease[32]
- Standardized assays and treatment regimens[32]
- Measurement of existing prognostic factors and confounders[32]
- Multivariable statistical analysis that includes the novel and existing prognostic factors[32,39]
- General guidance around the number of events required for adequate statistical power is 5 to 10 events per variable[40]

Once established from a reliable study, the clinical significance of a prognostic factor (i.e., the degree to which the factor might influence patient outcomes) needs to be carefully considered.[41] Before being accepted into standard practice, a prognostic factor's value should be considered alongside (1) reproducible and widely available test methods for the prognostic factor;

(2) predictions with therapeutic implications that can be interpreted by clinicians and benefit the patient; and (3) conclusions that are based on independently confirmed studies.[41]

Context-Relevant Seminal Studies of MRD

In the past few decades, a great deal of research has been conducted, and as a result MRD has strongly influenced treatment protocols. A considerable body of literature now exists to demonstrate that the overall prognosis (overall survival, EFS, relapse risk) differs for groups of MRD-positive patients compared with those who are MRD-negative at various time points during treatment: high-MRD patients with a high risk of relapse, and low-MRD patients with favourable prognosis.[20,26,27] The prognosis differential was also found in patients with relapsed ALL.[14] Local clinical experts advised that clinical practice in Ontario has been shaped by some key, seminal studies of MRD, summarized below.

De Novo ALL

A large study in 2008 from the Children's Oncology Group was influential and relevant to local practice. The results published by Borowitz et al[42] included a large sample of North American patients with de novo ALL and examined the relationship between MRD and other, established prognostic factors in ALL. At the time, prior studies of MRD were criticized for small size and methodological shortcomings; a large, multicentre study in the North American context was needed. In this study, more than 2,000 patients with precursor-B ALL (precursor B-cell) were enrolled between 2000 and 2005 and had MRD evaluated in bone marrow at the end of induction and again at the end of consolidation, as well as in the peripheral blood on Day 8 after the start of treatment. Minimal residual disease was evaluated via four-colour flow cytometry for all patients. The prognostic value of MRD for EFS and its relation to other prognostic factors—National Cancer Institute criteria, cytogenetics, treatment response—was analyzed. Events were defined as relapse, second malignancy, death, or last date of contact for EFS analyses. The MRD prognostic findings are summarized in Table 5.

Table 5: Summary of MRD Relation to EFS from a Large Children's Oncology Group Study

MRD Timing	Sample	MRD+ Cut-point	N Analyzed	Variables in Analysis[a]	HR for EFS
Day 8	PB	> 0.01%	1,946	End-induction MRD, NCI risk group, favorable trisomies, day-8 MRD, TEL-AML1 translocation, Day-8 M1 marrow	1.51[b]
End of induction	BM	> 0.01%	2,086		4.31[b]
End of consolidation	BM	> 0.01%	1,470	End-induction MRD, NCI risk group, End-consolidation MRD	2.25[b]

Abbreviations: BM, bone marrow; EFS, event-free survival; HR, hazard ratio; M1, bone marrow with < 5% leukemic blast cells; MRD, minimal residual disease; N, sample size; NCI, National Cancer Institute; PB, peripheral blood; TEL-AML1, favourable genetic translocation.
[a]Stepwise Cox regression analyses.
[b]Statistically significant at the level of $P < .001$.
Source: Borowitz et al, 2008.[42]

The above significant prognostic value of MRD for EFS in de novo cases shaped subsequent Children's Oncology Group ALL treatment protocols, including those used locally, which went on to include MRD evaluation at the end of induction to alter intensity of post-induction therapy.[42] These results were replicated in other landmark studies by other research groups.

In 2010, the Associazione Italiana Oncologia Pediatrica (AIEOP) and Berlin-Frankfurt-Münster (BFM) Study Group also published results from a large study (N = 3,184) to identify the strongest prognostic factor for pediatric ALL.[43] The AIEOP-BFM 2000 protocol evaluated MRD with PCR and created risk groups (low, intermediate, and high) on the basis of MRD information at end-induction (time point 1) in combination with MRD at end-consolidation (time point 2). The Cox analysis of 2,927 patients included other prognostic factors, namely age, white blood cell count, *TEL-AML1* status, DNA index, prednisone response, and MRD risk group.

After adjustment, those in the MRD–intermediate-risk group had a significantly higher likelihood of relapse, as did the MRD–high-risk group compared with MRD–standard-risk.[43] In an analysis of only patients with T-cell ALL from the same trial, MRD–high-risk patients had a significantly increased risk of relapse compared with the MRD–intermediate-risk patients, after adjustment for age, white blood cell count, prednisone response, and subtype of T-cell ALL.[44] The AIEOP-BFM groups have used the MRD time point 1 + time point 2 combination to define risk groups in many of their subsequent protocols and found a valuable ability to predict relapse, as well as EFS. Similar results have been found for infants with ALL using similar MRD risk groups (low, medium, high) on the basis of two analogous time points.[45] While this study does not answer the exact research question of this review because of the timing and categorization of MRD, this summary is provided because these studies are widely cited and relevant to practice patterns in Canada.

Relapsed ALL and Hematopoietic Stem Cell Transplant Recipients

The BFM research group studied the outcomes and prognostic value of MRD in relapsed ALL.[46] Patients with B-cell ALL were treated on the ALL-REZ-BFM 95/96 protocol, which aimed to develop a reliable way of identifying the relapsed intermediate-risk patients who require HSCT. Polymerase chain reaction was used to evaluate MRD at the end of re-induction therapy, and the occurrence of adverse events over 10 years was compared between MRD-negative (< 0.001%) and MRD-positive (≥ 0.001%) groups. Patients who were MRD-positive had both significantly worse EFS (defined as relapse, death after remission, or secondary malignancy) and higher cumulative incidence of relapse than those who were MRD-negative.[46] After forward stepwise testing, a final multivariable model showed MRD-positive status to be the only significant prognostic factor for EFS.[46]

In follow-up to ALL-REZ-BFM 95/96, the BFM collaboration studied the effect of systematic allocation to HSCT of these intermediate-risk relapsed patients classified as MRD-positive after re-induction. Considered the landmark study for the transplant cohort, 208 eligible patients at first relapse had MRD evaluated by PCR.[47] Patients were classified as molecular poor responders (MRD-positive ≥ 0.001) or molecular good responders (MRD-negative < 0.001) at the end of re-induction. Those who were MRD-positive were eligible for HSCT (83% underwent transplantation), and those who were MRD-negative received chemotherapy and irradiation. The study found that EFS was improved by HSCT in MRD-positive relapsed patients, as well as in MRD-negative patients via chemotherapy or irradiation. The BFM group validated the cut-point used in the study (MRD-positive ≥ 0.001%) for use as an indication for HSCT, reporting beneficial results in comparison with the preceding treatment protocol.[47] The Limited Institution Study for Treatment of Acute Lymphoblastic Leukemia for Paediatric Patients in Australia, New Zealand and The Netherlands (ANZCHOG ALL8) trial corroborated the BFM findings with their published results in 2015.[48] Among the 81 included patients transplanted once achieving first or subsequent complete remission, the prognostic value of MRD-positivity pre-HSCT (defined according to EuroMRD guidelines) was found to be significant for leukemia-free survival

(defined as relapse, death, or last follow-up).[48] This summary has been included because these studies are widely cited and relevant to practice patterns in Canada.

Context of Current Review

The compelling relation of MRD to EFS also informed subsequent research profoundly; treatment protocols and MRD studies have incorporated MRD evaluation and use in clinical practice. Not surprisingly, few RCTs have been published since these large international studies and local experts advised that clinicians are concerned that it is no longer ethical to randomize evaluation of MRD given the known difference in prognosis between MRD groups. Many published MRD studies therefore describe the overall prognosis of patients (report the differences in proportions for outcomes between MRD-positive and MRD-negative patients) in current clinical trials relative to predecessor treatment protocols. In this clinical area, broad adoption of trial protocols is the normative route through which standard practice is formed as drug regimens evolve.

The key seminal studies are very powerful because ultimately survival is the most important patient outcome. How patients feel and function and whether they survive (overall survival) are the most important and direct measures of clinical efficacy of an intervention.[49] Overall survival has long been considered the gold standard, "hard" outcome in oncology trials, in large part because of its precision, ease, and directness as a measure of benefit.[50] Though there is near-universal acceptance of overall survival as an outcome, deaths from causes other than cancer are included in overall survival, and large studies with long follow-up are required.[50] In leukemia specifically, duration of survival is not necessarily related to the specific measures of treatment success (e.g., achieving a short period of complete remission could exert little effect on overall survival).[51] Other clinical end points for leukemia clinical trials include disease-free survival, EFS, complete remission, improved end-organ function (e.g., restoration of normal hematopoiesis), and quality of life.[51]

The National Cancer Institute provides clinically meaningful definitions for EFS, disease-free survival, and relapse-free survival, but across studies definitions are heterogeneous in the signs and symptoms (events) accounted for. The trend away from overall survival toward other end points is warranted in order to expedite the availability of novel treatments to patients.[50] Among a number of benefits, faster time to trial completion and insights into mechanisms of action are key to treatment progress.[50] The primary goal of clinical research is to obtain information about whether there is clinically meaningful benefit, in a reliable way.[49] Validation of surrogate outcomes relative to overall survival is critical, and a strong correlation with an important clinical outcome alone is inadequate.[50] A valid surrogate end point must be associated with the ideal end point itself (overall survival), and be associated with the treatment; the treatment must also work through the surrogate end point to affect the ideal end point.[50] Even in the presence of valid surrogate end points, caution must be exerted when extrapolating from the results.[50]

Event-free survival is a widely used end point that can capture different clinically relevant complications and consequences. In MRD studies, EFS is often defined to include one or more of relapse, second malignancy, treatment-related toxic effects, mortality, morbidity, and drug resistance (Figure 3). Authors of MRD studies frequently describe the primary determinant of poor EFS as relapse, accounting for nearly all events in this composite outcome. Authors of several seminal studies were contacted and confirmed this was the case.[42,46] Treatment-related mortality, drug resistance, treatment toxicity, and second malignancy occur so infrequently that their contribution is likely to be very low. While the other events are relatively rare, EFS defined as multiple types of events provides only an indirect measure of relapse. The advantage of

examining relapse alone is the potential to quantify and characterize the relationship between MRD and EFS, providing further insight to aid clinical practice. This review sought to further elucidate the relationship by looking directly at relapse, the biological mechanism through which MRD is thought to act.
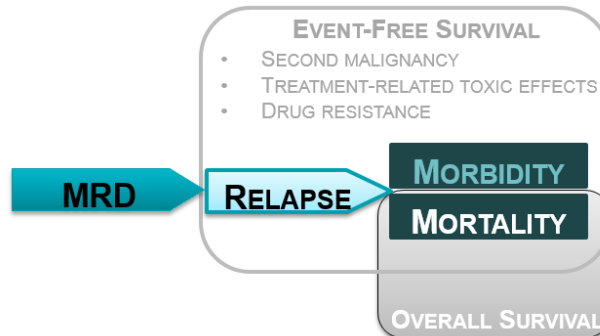
Figure 3: Relationship of Relapse to Common Research End Points in Leukemia Clinical Trials

Abbreviation: MRD, minimal residual disease.

Ontario Context

There are approximately 21 new cases per year of AML and 109 new cases per year of ALL in Ontario,[20] where all pediatric patients with ALL or AML enrolled in therapeutic studies currently have access to MRD evaluation funded by reference laboratories in the United States. The funding for MRD evaluation via therapeutic trial enrollment will end on June 1, 2016, and will be replaced with a fee-for-service model. Minimal residual disease testing via flow cytometry has been made available at one laboratory in Toronto for patients with (precursor) B-cell ALL who are not enrolled in studies. Children with T-cell ALL or AML who are not enrolled in therapeutic studies have access to MRD evaluation via out-of-province reference laboratories for which Ontario hospitals are billed on a fee-for-service basis. Currently for ALL, MRD is evaluated at baseline, Day 8 of treatment, after induction, and after consolidation. Because of low incidence, T-cell ALL and AML will likely continue to be conducted in out-of-country reference laboratories, whereas B-cell ALL will be evaluated in Ontario. The Provincial Oncology Group of Ontario (POGO) MRD Working Group, after initial cost investigation, speculated that it might be less expensive to test for MRD in Ontario than to send samples on a fee-for-service model to US reference laboratories. In 2014, the Foundation for the National Institutes of Health and partners including the National Cancer Institute launched an MRD study that includes undertaking standardization of flow cytometric measurement of MRD across North America, and will be coordinated with the Children's Oncology Group program to transition its MRD measurements to a new laboratory service model.[28] The POGO MRD Working Group is developing a plan for the implementation of MRD evaluation in Ontario, making use of the existing Toronto site as the primary local laboratory and including development of standardized flow cytometry protocols.

CLINICAL EVIDENCE REVIEW

Research Questions

1. **In children with ALL, is MRD an independent prognostic factor for disease relapse?** Specifically:
   a. In patients with de novo ALL, is an MRD-positive result at the end of induction therapy associated with increased risk of first relapse?
   b. In patients with de novo ALL, is an MRD-positive result at the end of consolidation therapy associated with increased risk of first relapse?
   c. In patients with relapsed ALL, is an MRD-positive result after re-induction therapy associated with increased risk of subsequent relapse?
   d. In patients with ALL who receive HSCT, is an immediate pretransplant MRD-positive result associated with increased risk of relapse after transplant?

2. **In children with ALL, is MRD predictive of benefit from therapy?**
   Specifically, in patients with ALL, what is the effect of MRD-based risk-directed treatment compared with clinical risk stratification for treatment selection on EFS, overall survival, and relapse?

Methods

Literature Search Strategy

A literature search was performed on November 3, 2014, using Ovid MEDLINE, Ovid Embase, Cochrane Central Register of Controlled Trials, Cochrane Database of Systematic Reviews, Database of Abstracts of Reviews of Effects (DARE), Health Technology Assessment Database and National Health Service (NHS) Economic Evaluation Database for studies published from January 1, 1998, to November 3, 2014. (Appendix 1 provides details of the search strategy.) Abstracts were screened by a single reviewer and, for those studies meeting the eligibility criteria, full-text articles were obtained. Reference lists were also examined for any additional relevant studies not identified through the search.

Grey literature sources used included clinical experts, ClinicalTrails.gov, the World Health Organization International Clinical Trials Registry Platform, and the Children's Oncology Group. Eligible grey literature was reported without data extraction, analysis, or quality assessment.

Inclusion Criteria

- English-language full-text publications
- Published between January 1, 1998, and November 3, 2014
- Pediatric (aged < 18) or adolescent and young adult (aged 19–30) study populations; or mixed-age study populations that include pediatric (aged < 18) or adolescent and young adult (aged 19–30) patients and report results by these discrete age categories
- Patients with ALL or mixed leukemia study populations that include patients with ALL and report results discretely for ALL
- Reports discrete data for ≥ 1 outcome of interest (i.e., **Question 1a–d** relapse; **Question 2** EFS, overall survival, and relapse)
- MRD evaluation (i.e., quantification of leukemic blasts in bone marrow) of all patients with either:

- o Polymerase chain reaction: DNA–based PCR methods on T-cell receptor gene, or immunoglobulin rearrangements; or
- o Flow cytometry: leukemia-associated immunophenotype or difference-from-normal
- For **Question 1a–d:**
  - o Prospective cohort studies (including those using historical data)
  - o Reporting results for ≥ 1 discrete population and ≥ 1 time point specified from multivariable regression analysis; if no multivariable estimates are available in the literature, estimates from univariate regression analysis
  - o For **Question 1d only:** HSCT recipients: allogenic stem cell transplant from any source (including peripheral blood, bone marrow, or umbilical cord blood)
- For **Question 2:**
  - o Randomized controlled trials (RCTs) (including cluster RCTs), prospective comparative cohort studies; or systematic reviews of the above study types
- Eligible clinical risk stratification comparators:
  - o **De Novo ALL**—National Cancer Institute criteria (patient age, white blood cell count, and CNS disease), cytogenetics, and early response to treatment (i.e., morphologic evaluation of bone marrow early or during induction therapy, or prednisone response); or study-specific risk-stratification criteria
  - o **Patients with ALL who receive HSCT**—Site of relapse, time interval to relapse from diagnosis, leukemic cell lineage (B-cell vs. T-cell)

Exclusion Criteria

- Non-English full-text publications, conference abstracts, editorials, presentations, unpublished studies, duplicate reports
- Noncomparative or historical-controlled cohort studies (**Question 2**), cross-sectional or case-control studies, time series, pre-post study designs, case reports, or case series
- Studies without an eligible comparator for MRD-directed treatment
- Studies on adults > 30 years old, or with mixed-age populations without discrete reporting of results for relevant age groups
- Studies on patients with other leukemias or hematological malignancies
- Studies on overall prognosis or prognostic factors other than MRD, or on the feasibility or validation of novel markers for MRD
- Studies using other techniques for MRD evaluation (e.g., PCR-fusion transcripts)
- Studies using differing MRD evaluation techniques on subsets of patients within a study (e.g., real-time quantitative PCR on 77%, dot-blot hybridization on 8%, both on 15%)
- Analysis with other single or combined time points for prognostic evaluation (e.g., after early intensification, or mid-induction + post-induction MRD result combined)
- Those not reporting discrete data for outcome(s) of interest, or composite or undefined end points (e.g., EFS defined as relapse, second malignancy, or death)

Research Question 1

- Relapse (including EFS, where relapse is the only event)

Research Question 2

- Overall survival
- Event-free survival
- Relapse (e.g., time-to-relapse, relapse-free interval)

Statistical Analysis

Adjusted hazard ratio and variance (log[HR]) were sought to enable pooling data from prognostic studies' multivariable proportional hazards model. Authors were contacted via email where there were missing or incomplete data reported, or where clarification was needed regarding outcome definition or the published analysis. Meta-analysis was inappropriate because of statistical and methodological heterogeneity. RevMan 5.2 was used for figure generation.

Statistical Heterogeneity and Effect Modifiers

Heterogeneity precluded pooling data; thus, a priori subgroup analyses by age, type of assay, study design, and risk of bias could not be performed.

Publication Bias

Publication bias could not be assessed using funnel plot methodology or statistical tests (e.g., Egger's, Begg's) because the number of studies was insufficient.

Quality of Evidence

For risk-directed therapy, the quality of the body of evidence for each outcome was examined according to the Grading of Recommendations Assessment, Development, and Evaluation (GRADE) Working Group criteria.[52] The overall quality was determined to be high, moderate, low, or very low using a step-wise, structured method.

Study design was the first consideration; for intervention studies the starting assumption was that RCTs are high quality, whereas observational studies are low quality. Five additional factors—risk of bias, inconsistency, indirectness, imprecision, and publication bias—were then taken into account. Limitations in these areas resulted in downgrading the quality of evidence. Finally, three main factors that can raise the quality of evidence were considered: the large magnitude of effect, the dose-response gradient, and any residual confounding factors.[52] For more detailed information, please refer to the latest series of GRADE articles.[52]

The GRADE Working Group adapted considerations of the above GRADE domains.[53] This GRADE guidance was used for evaluating the quality of prognostic studies. Individual studies were assessed for risk of bias using the Quality In Prognosis Studies tool (QUIPS).[54] For intervention studies, the GRADE risk of bias considerations were used.

As stated by the GRADE Working Group, the final quality score can be interpreted using the following definitions:

High	High confidence in the effect estimate—the true effect lies close to the estimate of the effect
Moderate	Moderate confidence in the effect estimate—the true effect is likely to be close to the estimate of the effect, but may be substantially different
Low	Low confidence in the effect estimate—the true effect may be substantially different from the estimate of the effect
Very Low	Very low confidence in the effect estimate—the true effect is likely to be substantially different from the estimate of the effect

Findings of Evidence Review

The database search yielded 2,832 citations published between January 1, 1998, and November 3, 2014 (with duplicates removed). Articles were excluded on the basis of information in the title and abstract. The full texts of potentially relevant articles were obtained for further assessment. Figure 4 shows the breakdown of when and for what reason full-text citations were excluded from the analysis.

Nine authors were contacted for clarification on one or more of the following in order to determine eligibility: outcome definition, analysis or effect estimates, and parameters. Four responded with additional information that facilitated inclusion[55-58]; five studies were excluded because clarifying information showed ineligibility[31,42,46] or because of non-response from authors.[59,60]

Twelve studies (11 prospective cohort studies and one RCT) reported in 14 articles met the inclusion criteria.[55-58,61-70] Reference lists of the included studies and health technology assessment websites were searched to identify other relevant studies, and no additional citations were included.
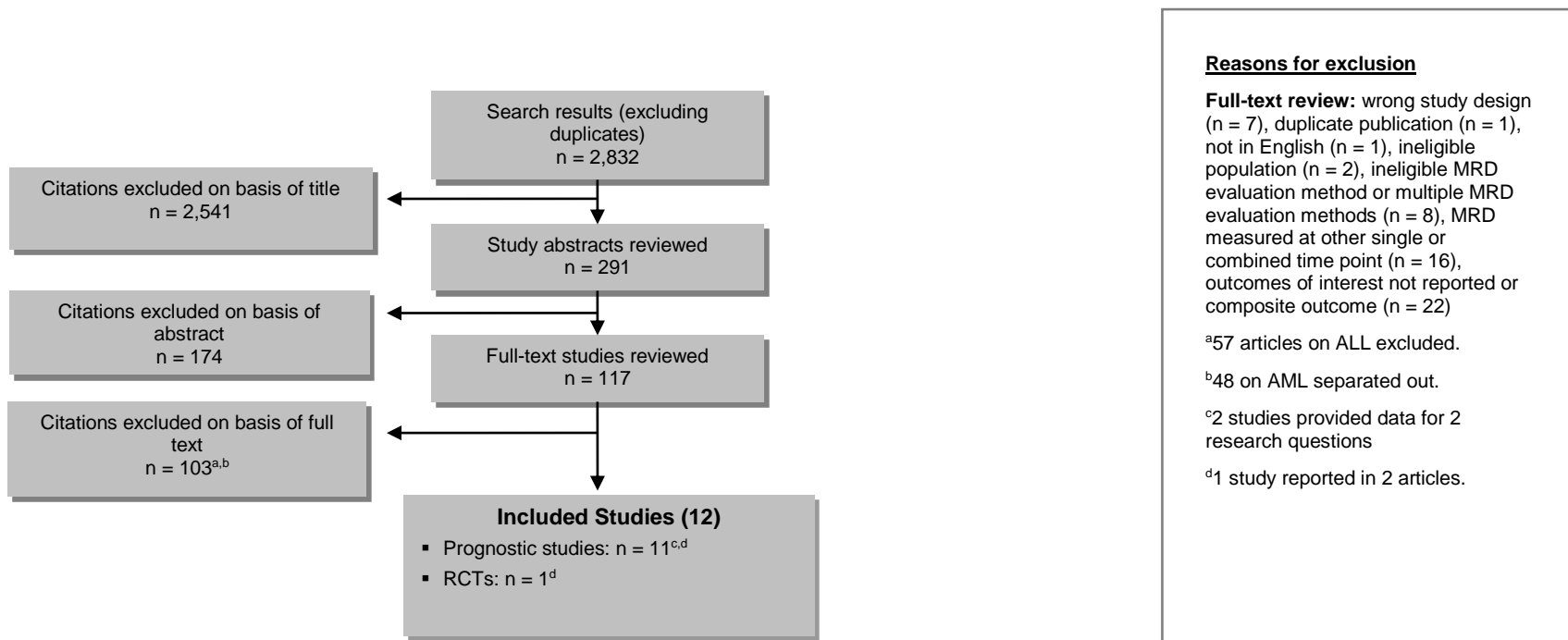
Figure 4: Citation Flow Chart

Abbreviations: ALL, acute lymphoblastic leukemia; AML, acute myeloid leukemia; MRD, minimal residual disease; RCT, randomized controlled trial.

Question 1: Prognostic Studies

Question 1a: Patients With De Novo ALL With MRD Measured at End of Induction

A total of seven studies reported in eight articles evaluating the independent prognostic significance for relapse of MRD evaluation at the end of induction were identified.[56-58,62,64,66,67,70] Six were prospective cohort studies, one used prospective historical data,[66] and one used some historical prospective and some real-time cohort data.[67] The studies ranged from single-site to multicentre studies recruiting patients between the early 1980s to late 2000s in Europe, North America, and Australia. Table 6 summarizes the characteristics of these studies.

Table 6: Characteristics of Prognostic Studies Evaluating MRD at End of Induction

Author, Year	Country	Time Period	MRD Sample Size (Total)	ALL Type, %	Median Age, Years (Range)	Male, %	MRD Assay (Target)
Meleshko et al, 2011[67]	Belarus	2003–2009	68 (279)	B-cell 78.0 T-cell 22.0	6.4 (1.4–22.8)	55.9	RQ-PCR (Ig/TCR)
Zhou et al, 2007[70]	United States, Canada	1996–2000	284 (498)	B-cell	NR	51.4	RQ-PCR (Ig/TCR)
Laughton et al, 2005[66]	Australia	1986–1998	62 (227)	B-cell 87.0 T-cell 10.0 Unknown 3.0	3.8 (1.3–15.9)	38.1	Real-time PCR (Ig/TCR)
Dworzak et al, 2002[64]	Austria	1996–1998	108 (139)	B-cell 92.6 T-cell 7.4	3.8 (0.08–17.08)	58.3	FCM (LAIP)
Van Dongen et al, 1998[58]	Austria, Germany, Italy, Netherlands	1991–1995	240 (625)	B-cell 87.5 T-cell 12.5	NR	NR	PCR (Ig/TCR)
Cavé et al, 1998[56,a]	Belgium, France, Portugal	1989–1996[b]	178 (246)	B-cell 84 T-cell 16	NR[c]	NR	PCR (Ig/TCR)
Coustan-Smith et al, 1998, 2000[57,62,a]	United States	1991–1998	165 (195)	B-cell 71.5 T-cell 28.5	NR[d]	64.2	FCM (LAIP)

Abbreviations: FCM, flow cytometry; Ig/TCR, immunoglobulin heavy chain or T-cell receptor gene rearrangements; LAIP, leukemia-associated immunophenotype; MRD, minimal residual disease; NR, not reported; PCR, polymerase chain reaction; RQ-PCR, real-time quantitative PCR.
[a]Qualitative description of study only.
[b]Recruitment at 1 centre from 1989 to 1996, and 10 additional centres from 1993 to 1996.
[c]Median age and range not reported; however, participant age breakdown was as follows: 6% 0–1 year; 81% 2–9 years; 13% 10–15 years.
[d]Median age and range not reported; however, participant age breakdown was as follows: 3.6% < 1 year; 63.0% 1–9; 33.4% >9 years.

Studies of MRD are limited by the challenge of identifying reliable targets for a laboratory test that results in smaller samples of patients analyzed for MRD than of the study overall (Table 6). Most studies used PCR techniques for MRD evaluation; only two used flow cytometry.[57,62,64,66] The quantitative range (i.e., lowest sensitivity of detection) of the assays in the studies were all on the order of 10^{-4} (i.e., 1 in 10,000 cells), with the minimum detection of 10^{-3} (i.e., 1 in 1,000 cells). Inclusion criteria across studies were pediatric patients with de novo ALL enrolled in the various treatment protocols, and of course a bone marrow sample available and suitable for MRD evaluation. Treatment protocols varied across studies and as a product of the evolution of ALL treatment over time. In general, induction therapy consisted of 4 to 5 weeks of multiagent chemotherapy. In one study induction was preceded by a separate steroid pretreatment[56] and in another case was composed of two successive phases of induction chemotherapy.[64]

All studies statistically analyzed the prognostic value of MRD after induction therapy on relapse risk, controlling for one or more known confounders (i.e., established prognostic factors; range 1–6). There was significant heterogeneity, both clinically and methodologically, that made quantitative synthesis inappropriate. Studies were inconsistent in the MRD cut-points or categorization, statistical techniques, and variables adjusted for in the analyses (Table 7).

Table 7: Summary of Findings of Prognostic Significance of MRD at End of Induction

Author, Year	MRD Assay	MRD Categorization, % (% of participants)	Variables in Analysis	Relapse Effect Estimate (95% CI)	*P*
Meleshko et al, 2011[67]	RQ-PCR	MRD+: ≥ 0.0001 (74.6) MRD−: < 0.0001 (25.6)	MRD, WBC count at diagnosis, immunophenotype, size of spleen, sex	HR 2.41 (NR)	.005
Zhou et al, 2007[70]	RQ-PCR	High: ≥ 0.001 (86.6) Low: < 0.001 (13.4)	MRD, risk group,[a] treatment group	HR 10.6 (6.05–18.55)	.001
Laughton et al, 2005[66]	Real-time PCR	High: ≥ 0.001 (69.4) Low: < 0.001 (30.6)	MRD, risk group,[b] end-induction WBC, end induction Absolute neutrophil count	HR 3.03 (1.20–7.65)	.019
Dworzak et al, 2002[64]	FCM	MRD+: ≥ 0.01 (40.9) MRD−: < 0.01 (59.1)	MRD risk at Week 12, BFM risk group[c]	HR 14.9 (3.2–66.7)	< .001
Van Dongen et al, 1998[58]	PCR	High: ≥ 0.01 (7.9) Intermediate: 0.001 (71.2) Low: ≤ 0.0001 (20.9)	MRD, treatment group,[d] age, sex, immunophenotype, WBC (continuous), country, MRD-treatment group interaction	Relative relapse rate per 10-fold increase in MRD 1.6 (1.3–.9)	< .001
Cavé et al, 1998[56,e]	PCR	MRD+: ≥ 0.01 MRD−: < 0.01	MRD, immunophenotype	HR 10.6[f] (3.9–28.7)	NR
Coustan-Smith et al, 1998, 2000[57,62,g]	FCM	MRD+: ≥ 0.01 (74.5) MRD−: < 0.01 (25.5)	MRD, age	NR	< .001
			MRD, WBC	NR	< .001
			MRD, Ph+	NR	< .001
			MRD, *MLL*	NR	< .001
			MRD, either Ph+ or *MLL*	NR	.004

Abbreviations: BFM, Berlin-Frankfurt-Münster; CI, confidence interval; FCM, flow cytometry; HR, hazard ratio; *MLL*, adverse genetic feature; MRD, minimal residual disease; NR, not reported; PCR, polymerase chain reaction; RQ-PCR, real-time quantitative PCR; Ph+, Philadelphia chromosome–positive adverse genetic feature; WBC, white blood cell.

[a]Risk group defined as standard- or high-risk on basis of age, WBC, immunophenotype, central nervous system involvement, and mediastinal mass. Treatment group defines treatment on basis of randomization to 1 of 2 asparaginases, or direct-assigned versus randomized to *Escherichia coli* treatment.
[b]Risk group defined as standard or high risk on basis of National Cancer Institute criteria: age and WBC.
[c]BFM risk group defined as standard-, medium-, or high-risk on basis of prednisone response, WBC, immunophenotype, age, and cytogenetics.
[d]Treatment group defined by presenting leukemic cell mass and prednisone response. Results presented for Time Point 2.
[e]Results of stratified Cox model presented.
[f]Analysis of present (≥ 1.5×10^{-4}) versus absent (< 1.5×10^{-4}) MRD was also conducted and yielded similar results (HR = 5.3, 95% CI 2.2–12.6).
[g]Descriptive information available only for MRD results.

Few studies reported data for the multivariable analysis in detail; however, all studies reported a statistically significant association between the presence or level of MRD after induction and elevated likelihood of relapse. Cavé et al[56] reported using a Cox model to determine the most significant independent prognostic factors and finding MRD was the most important prognostic factor followed by immunophenotype or white blood cell count. However, data for this analysis were not published, so results of the Cox analysis stratified by only immunophenotype are presented in Table 7 (above), and Table 8 for end of consolidation.

One study[57,62] conducted a series of separate, adjusted analyses but reported only *P* values. Thus, this study could be evaluated for risk of bias only (Table A1) and was not included in the GRADE assessment. A single study reported an estimate reflecting continuous MRD levels[58] and found a positive linear relationship of 60% higher risk of relapse for every 10-fold increase in MRD (Table 7).

The finding per 10-fold increase in level is a piece of evidence suggesting a relationship between MRD threshold and relapse risk. Plotting the effect estimates of the other five studies by cut-point did not reveal a distinct pattern (data not shown). There is much uncertainty because of the small number of events across studies; consequently, the possibility of a relationship cannot be ruled out. The use of differing cut-points for MRD-positive and MRD-negative status in the analyses and significant unaccounted-for heterogeneity precluded meta-analysis of adjusted hazard ratios. The adjusted hazard ratios reported from all studies evaluating MRD at the end of induction are presented visually only in Figure 5 (values might differ slightly from those in Table 7 because of rounding errors).

Study or Subgroup	log[Hazard Ratio]	SE	Hazard Ratio IV, Random, 95% CI	Year	Hazard Ratio IV, Random, 95% CI
Cavé 1998 (1)	2.3609	0.5082	10.60 [3.92, 28.70]	1998	
Dworzak 2002 (2)	2.7014	0.7777	14.90 [3.25, 68.42]	2002	
Laughton 2005 (3)	1.1086	0.4726	3.03 [1.20, 7.65]	2005	
Zhou 2007 (4)	2.3609	0.2861	10.60 [6.05, 18.57]	2007	
Meleshko 2011 (5)	0.8796	0.3134	2.41 [1.30, 4.45]	2011	

0.02   0.1   1   10   50
Reduced risk of relapse   Elevated risk of relapse

(1) Stratified by immunophenotype
(2) Adjusted for risk group based on prednisone response, WBC, immunophenotype, age, and cytogenetics
(3) Adjusted for NCI risk group, end-induction WBC, end induction absolute neutrophil count
(4) Adjusted for risk group based on age, WBC, immunophenotype, CNS, and mediastinal mass, and treatment group
(5) Adjusted for WBC count at diagnosis, immunophenotype, size of spleen, and sex

Figure 5: Adjusted Hazard Ratios for Relapse Associated With MRD+ Status at End-Induction in De Novo ALL

Abbreviations: CI, confidence interval; IV, inverse variance; MRD, minimal residual disease; NCI, National Cancer Institute; SE, standard error; WBC, white blood cell.

There was a trend toward more modest effect sizes and precise point estimates when more variables were adjusted for in the analysis (e.g., van Dongen et al[58] and Laughton et al[66]). The relapse rates in these studies ranged from 12.0% to 32.3% and might influence the width of confidence intervals, especially in those studies with lower rates (e.g., Dworzak et al[64]). Only one study accounted for adverse cytogenetic features in their analysis,[64] without which residual confounding could have influenced results. Direct comparison of the point estimates across studies is hampered by differing methods. Regardless of the MRD categorization, cut-point, or use of flow cytometry or PCR assays, a positive MRD result at the end of induction was

associated with significantly increased risk of relapse, after accounting for one or more established prognostic factors. The GRADE quality of evidence for this body of evidence is Low (full assessment is presented in Table A2).

Question 1b: Patients With De Novo ALL With MRD Evaluated at End of Consolidation

Two studies that examined end-induction MRD also separately evaluated the prognostic value of MRD at the end of consolidation therapy.[56,58] Both studies were conducted in Europe in the late 1980s to mid-1990s and used PCR techniques (Table 8).

Table 8: Characteristics of Prognostic Studies Evaluating MRD at End of Consolidation

Author, Year	Country	Time Period	MRD Sample Size (Total)	ALL Type(s)	Median Age, Years (Range)	Male (%)	MRD Assay (Target)
Van Dongen et al, 1998[58]	Austria, Germany, Italy, Netherlands	1991–1995	240 (625)	B-cell 87.5% T-cell 12.5%	NR	NR	PCR (Ig/TCR)
Cavé et al, 1998[56]	Belgium, France, Portugal	1989–1996	162[a] (246)	B-cell 84% T-cell 16%	NR[b]	NR	PCR (Ig/TCR)

Abbreviations: ALL, acute lymphoblastic leukemia; Ig/TCR, immunoglobulin heavy chain or T-cell receptor gene rearrangements; MRD, minimal residual disease; NR, not reported; PCR, polymerase chain reaction.
[a]Only standard-risk patients were analyzed at end of consolidation.
[b]Median age and range not reported; however, participant age breakdown was as follows: 6% 0–1 year; 81% 2–9 years; 13% 10–15 years.

Eligible patients had de novo ALL, presented at research sites, and could be evaluated for MRD because samples were available and PCR assay had adequate reproducibility. The quantitative range in these studies varied from 1.5×10^{-4} (Cavé et al[56]) to 1.5×10^{-6} (Van Dongen et al[58]), and follow-up ranged from 38 months[56] to 48 months.[58] In the study by Cavé et al, only the standard-risk patients enrolled in the study were assessed for MRD at the end of consolidation. Consolidation therapy took place over approximately 8 weeks for the European Organization for Research and Treatment of Cancer (EORTC) 58881 and 4 weeks for the BFM 90 protocols.[56,58]

Table 9: Summary of Findings of Prognostic Significance of MRD at End of Consolidation

Author, Year	MRD Assay	MRD Categorization, % (% of participants)	Variables in Analysis	Relapse Effect Estimate (95% CI)	P
Van Dongen et al, 1998[58]	PCR	High: ≥ 0.01 (8.9) Intermediate: 0.001 (66.2) Low: ≤ 0.0001 (24.9)	MRD, treatment group,[a] age, sex, immunophenotype, WBC (continuous), country, MRD-treatment group interaction	Relative relapse rate per 10-fold increase in MRD 1.9 (1.6–2.2)	< .001
Cavé et al, 1998[56,b]	PCR	MRD+: ≥ 0.001 MRD−: < 0.001	MRD, immunophenotype	HR 11.2[c] (3.6–34.7)	NR

Abbreviations: CI, confidence interval; HR, hazard ratio; MRD, minimal residual disease; NR, not reported; PCR, polymerase chain reaction; WBC, white blood cell.
[a]Treatment group defined by presenting leukemic cell mass and prednisone response. Results presented for Time Point 3 (data provided by authors).
[b]Results of stratified Cox model presented.
[c]Analysis of present (≥ 1.5×10^{-4}) versus absent (< 1.5×10^{-4}) MRD was also conducted and yielded significant results (HR = 6.1, 95% CI 2.5–14.9).

Similar to the findings at end of induction, methodologic heterogeneity precluded quantitative synthesis. Van Dongen et al[58] evaluated MRD as a continuous variable (90% increased risk of relapse for every 10-fold increase in MRD level), whereas Cavé et al[56] categorized patients into MRD-positive and MRD-negative on the basis of a cut-point of ≥ 0.001 (Figure 6; values could differ slightly from those in Table 9 because of rounding errors). These differences preclude direct comparison of the point estimates and values.

Study or Subgroup	log[Hazard Ratio]	SE	Hazard Ratio IV, Random, 95% CI	Year	Hazard Ratio IV, Random, 95% CI
Cavé 1998	2.4137	0.5781	11.18 [3.60, 34.70]	1998	

Reduced risk of relapse   Elevated risk of relapse
0.05   0.2   1   5   20

Figure 6: Adjusted Hazard Ratio for Relapse Associated With MRD+ Status at End-Consolidation in De Novo ALL

Abbreviations: CI, confidence interval; IV, inverse variance; MRD, minimal residual disease; SE, standard error.

Both analyses yielded a statistically significant increase in risk of relapse among those with present or higher MRD at the end of consolidation. The relapse rates in the Cavé et al[56,58] and van Dongen et al[56,58] studies were 21.3% and 25%, respectively. Regardless of research methods, a positive MRD result at the end of consolidation was associated with significantly increased risk of relapse, after accounting for one or more established prognostic factors. The GRADE quality of evidence is Moderate and the full assessment can be found in Table A2.

Question 1c: Patients With Relapsed ALL With MRD Evaluated at End of Re-induction

A single study assessed the prognostic value of MRD for relapse after re-induction therapy, in patients with relapsed ALL.[63] This study was a prospective cohort enrolled in a treatment protocol for relapsed ALL. An overview of the study characteristics is presented in Table 10.

Table 10: Characteristics of Prognostic Studies Evaluating MRD After Re-induction Therapy in Patients With Relapsed ALL

Author, Year	Country	Time Period	MRD Sample Size (Total)	ALL Type, %	Median Age, Years (Range)	Male, %	MRD Assay (Target)
Coustan-Smith et al, 2004[63]	United States	1993–2003	35 (57)	B-cell 77.1 T-cell 22.9	10 (0.67–21)	68.6	FCM (LAIP)

Abbreviations: ALL, acute lymphoblastic leukemia; FCM, flow cytometry; LAIP, leukemia-associated immunophenotype; MRD, minimal residual disease.

A total of 57 participants in first relapse were enrolled between 1993 and 2003, of which 35 participants had available bone marrow samples and suitable leukemia-associated immunophenotypes for flow cytometric evaluation of MRD following re-induction therapy. Participants underwent 36-day re-induction therapy (according to St. Jude's R11 and R15 protocols) after which bone marrow samples were analyzed by three-colour (to August 1998) or four-colour (after August 1998) flow cytometry. Participants were categorized as either MRD-positive (i.e., ≥ 0.01% leukemic cells) or MRD-negative (i.e., < 0.01%), which represented 54.3% and 45.7% of the study sample, respectively.[63]

Statistical analysis included a multivariable proportional hazards model in which variables with a statistically significant univariate association with risk of second relapse ($P < .05$) were then combined. Potential covariates included time of relapse (on- or off-therapy), site of relapse (bone marrow, extramedullary, or combined), cell lineage (B- or T-cell), and percentage of circulating leukemic cells upon relapse ($< 10 \times 10^9$/L vs $\geq 10 \times 10^9$/L). Only relapse on-therapy and MRD+ status were significant in univariate analysis ($P = .001$ and $P = .02$, respectively) and so were included together in the multivariable analysis. The adjusted hazard ratio for second relapse after adjustment for on-therapy relapse was statistically significant at 3.40 (95% confidence interval [CI] 1.40–8.23, $P = .007$).[63] Values could differ slightly because of rounding errors (Figure 7).

Study or Subgroup	log[Hazard Ratio]	SE	Hazard Ratio IV, Random, 95% CI	Hazard Ratio IV, Random, 95% CI
Coustan-Smith 2004	1.2224	0.452	3.40 [1.40, 8.23]	

0.1 0.2   0.5   1   2   5   10
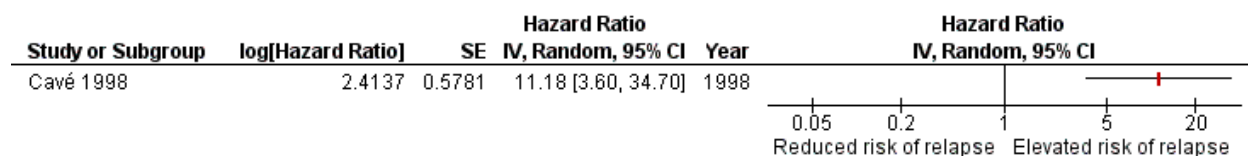Reduced risk of relapse   Elevated risk of relapse
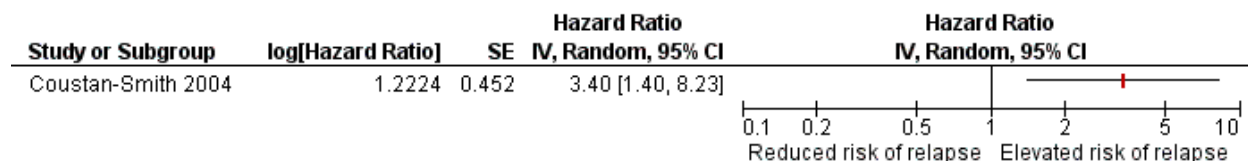
Figure 7: Adjusted Hazard Ratio for Relapse Associated With MRD+ Status at End of Re-induction in Relapsed ALL

Abbreviations: CI, confidence interval; IV, inverse variance; MRD, minimal residual disease; SE, standard error.

The sample size was very small with approximately 40% of patients experiencing second relapse during the study period.[63] This study was well conducted and showed that a positive MRD result after re-induction therapy in relapsed patients was associated with significantly increased risk of relapse, after accounting for time of relapse. The GRADE quality of evidence is Moderate (Table A2).

Question 1d: Patients With ALL Receiving HSCT With MRD Evaluated Before Transplantation

Three studies assessed the independent prognostic value of MRD evaluation before HSCT in patients with ALL.[55,61,65] All studies were conducted in European countries with patients in first or subsequent complete remission who received allogenic HSCT, where MRD was evaluated before transplantation and pre-transplant conditioning treatment. Table 11 outlines characteristics of the studies.

Table 11: Characteristics of Prognostic Studies Evaluating MRD Before HSCT

Author, Year	Country	Time Period	MRD Sample Size (Total)	ALL Type, %	Median Age, Years (Range)	Male (%)	MRD Assay (Target)
Gandemer et al, 2014[65]	France	2005–2008	122 (215[a])	High risk[b] or relapsed B-cell 70.5 T-cell 27.9	NR (6.9–7.7)	64.8	PCR (Ig/TCR)
Balduzzi et al, 2014[61]	Italy	2001–2011	82 (97)	First allogenic HSCT B-cell 85.4[c] T-cell 14.6	8 (< 1–20)	66.0	RQ-PCR (Ig/TCR)
Bader et al, 2002[55]	Germany	1986–1999	45 (59)	B-cell 84 T-cell 16	9.8 (1.5–17.8)	53.3	PCR (Ig/TCR)

Abbreviations: ALL, acute lymphoblastic leukemia; HSCT, hematopoietic stem cell transplant; Ig/TCR, immunoglobulin heavy chain or T-cell receptor gene rearrangements; MRD, minimal residual disease; NR, not reported; PCR, polymerase chain reaction; RQ-PCR, real-time quantitative PCR.
[a]Study enrolled patients with both ALL (n = 133) and myeloid leukemia (n = 82) and reported on ALL alone. Eleven of the patients with ALL could not be categorized according to pre-HSCT MRD and were not included in the analysis.
[b]High risk defined by unfavourable cytogenetics, induction failure, white blood cell count, and poor early response to treatment. Relapsed defined as early or very early bone marrow relapses.
[c]Calculated from data reported in article.

The studies generally included high-risk ALL, including relapsed patients and those who received their first allogenic HSCT. The sources of stem cells included bone marrow, peripheral blood, and cord blood. The important characteristics of the donors that influence the likely success of transplant (e.g., sibling donor or matched) and graft-versus-host disease (GVHD) were also included in analyses to account for the complication of HSCT whereby the transplanted cells attack the recipient's body.

One study[55] conducted MRD analysis but reported only a significant *P* value (Table 12); thus, this study could be evaluated for risk of bias only (Table A1) and was not included in the GRADE assessment. As no studies used the same definitions of MRD+ nor adjusted for the same variables, pooling the adjusted hazard ratios was inappropriate. The adjusted hazard ratios reported from the studies that provided them are presented visually only in Figure 8 (values could differ slightly from those in Table 12 because of rounding errors).

Table 12: Summary of Findings of Prognostic Significance of MRD Before HSCT

Author, Year	MRD Assay	MRD Categorization, % (% of participants)	Variables in Analysis	Relapse Effect Estimate (95% CI)	P
Gandemer et al, 2014[65]	PCR	MRD+: ≥ 0.001 (22.1) MRD−: < 0.001 (77.8)	MRD+/−, ATG, sex match vs F/M, M/F vs F/M, CNS location, IntReALL SR[a]	HR 3.932 (1.121–13.788)	.035
Balduzzi et al, 2014[61]	RQ-PCR	MRD+: ≥ 0.0001 (31.7) MRD−: < 0.0001 (68.3)	MRD+/−, disease phase, type of donor, HLA-compatibility, GVHD	HR 9.2 (3.54–23.88)	< .001
Bader et al, 2002[55]	PCR	MRD+: ≥ 0.001 MRD−: < 0.001	MRD+/−, pre-transplant CR status, type of donor (related vs. unrelated), sex, immunophenotype, acute GVHD disease	NR	.0095

Abbreviations: ALL, acute lymphoblastic leukemia; ATG, antithymocyte globulins (GVHD prophylaxis); CI, confidence interval; CNS, central nervous system; CR, complete remission; GVHD, graft-versus-host disease; HLA, human leukocyte antigen; HR, hazard ratio; HSCT, hematopoietic stem cell transplant; MRD, minimal residual disease; NR, not reported; PCR, polymerase chain reaction; RQ-PCR, real-time quantitative PCR; SR, standard-risk.

[a]IntReALL standard-risk defined according to the International Study for Treatment of Childhood Relapsed ALL 2010 (late isolated or late/early combined bone marrow relapse, any late/early isolated extramedullary relapse).
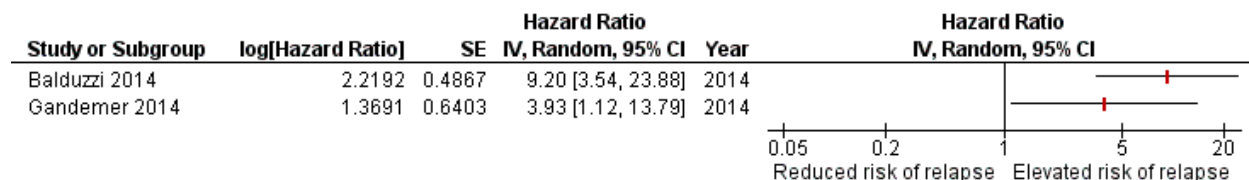
Figure 8: Adjusted Hazard Ratios for Relapse Associated With MRD+ Status Before Transplant in ALL Hematopoietic Stem Cell Transplant Recipients

Abbreviations: CI, confidence interval; IV, inverse variance; MRD, minimal residual disease; SE, standard error.

These data suggest an increase in likelihood of relapse after HSCT among MRD+ patients ranging a considerable spread of magnitude. Knowledge of this risk is valuable in clinical practice and could prompt preventive action; however, the action that can be taken carries its own risk. In the context before HSCT, the threshold of clinical significance (to motivate action) has yet to be conclusively defined, but would generally be very high because of the nature of the sequelae. A positive MRD result before transplant in HSCT recipients was associated with significantly increased risk of relapse, after accounting for donor characteristics, disease characteristics, and transplant complications. The GRADE quality of evidence is Moderate (Table A2).

Summary of Minimal Residual Disease Prognostic Studies for Relapse

Meta-analysis was inappropriate owing to methodological heterogeneity. Thus, the magnitude and precision of the prognostic effect of MRD at various time points during treatment could not be characterized. Despite heterogeneity in research methods (sample size, MRD cut-points, statistical analysis, and confounders) used in adjusted analyses, a positive MRD result in

patients with ALL was found to be a significant, independent prognostic factor for relapse when measured in each of the scenarios investigated in Questions 1a to d.

The threshold for clinical significance of the point estimates within the body of evidence was considered to be met overall. Experts advise that even a 5% increase in relapse risk is clinically significant, and all point estimates surpass that threshold. However, each scenario addressed in Question 1 has its own unique considerations depending on the preventive action that would be undertaken. For instance, where a clinical action is associated with lower risk (e.g., intensification of chemotherapy), the threshold of clinical significance is lower relative to when the clinical action is associated with high risk and long-term consequences (e.g., HSCT). Figure 9 provides a visual summary of the prognostic studies above, separated by research question and flow cytometry versus PCR (values could differ slightly from those reported above because of rounding errors).

Study or Subgroup	log[Hazard Ratio]	SE	Hazard Ratio IV, Random, 95% CI	Year	Hazard Ratio IV, Random, 95% CI
6.1.1 End of Induction - PCR					
Cavé 1998	2.3609	0.5082	10.60 [3.92, 28.70]	1998	
Laughton 2005	1.1086	0.4726	3.03 [1.20, 7.65]	2005	
Zhou 2007	2.3609	0.2861	10.60 [6.05, 18.57]	2007	
Meleshko 2011	0.8796	0.3134	2.41 [1.30, 4.45]	2011	
6.1.2 End of Induction - FCM					
Dworzak 2002	2.7014	0.7777	14.90 [3.25, 68.42]	2002	
6.1.3 End of Consolidation - PCR					
Cavé 1998	2.4137	0.5781	11.18 [3.60, 34.70]	1998	
6.1.4 End of Re-induction - FCM					
Coustan-Smith 2004	1.2224	0.452	3.40 [1.40, 8.23]	2004	
6.1.5 Pre-HSCT - PCR					
Gandemer 2014	1.3691	0.6403	3.93 [1.12, 13.79]	2014	
Balduzzi 2014	2.2192	0.4867	9.20 [3.54, 23.88]	2014	

0.02 0.1 1 10 50
Reduced risk of relapse   Elevated risk of relapse

Figure 9: Adjusted Hazard Ratios for Relapse Associated With MRD+ Status in Various Scenarios Across Studies of MRD Evaluation Using PCR and FCM

Abbreviations: CI, confidence interval; FCM, flow cytometry; HSCT, hematopoietic stem cell transplant; IV, inverse variance; MRD, minimal residual disease; PCR, polymerase chain reaction; SE, standard error.

Question 2: Studies of MRD-Directed Treatment

One RCT reported in two articles compared the effectiveness of MRD-directed treatment with standard treatment in groups defined by MRD status.[68,69] The UKALL2003 trial enrolled 3,126 eligible patients with ALL over a decade ending June 2011, of which 2,721 (87%) were determined to be at clinically standard or intermediate risk on the basis of National Cancer Institute criteria, cytogenetics, and early response to treatment (bone marrow morphology at Days 8 and 15). High-risk patients (Ph+ and mature B-cell leukemia) were ineligible, and patients with Down syndrome were not randomized because excess treatment-related mortality was identified mid-trial.[69]

Minimal residual disease was evaluated in clinically standard- and intermediate-risk patients who achieved complete remission after induction, separating them into two MRD groups (Table

13). Patients with MRD ≥ 0.01% at the end of induction therapy (Day 29) were classified as MRD–high-risk (MRD-HR), whereas MRD–low-risk (MRD-LR) patients were those with undetectable MRD at Day 29, as were patients who had MRD < 0.01% at Day 29 but undetectable MRD before the start of maintenance therapy. MRD-indeterminate patients had no or poor-quality bone marrow samples, or MRD < 0.01% that persisted before the start of maintenance therapy, and were ineligible for the study.[69]

Table 13: MRD Risk Groups of 2,721 Eligible Clinically Standard- and Intermediate-Risk Patients in UKALL2003 Trial

Countries	Time Period	MRD–High-Risk (%)	MRD–Low-Risk (%)	MRD Indeterminate (%)	Excluded[a] (%)	MRD Assay (Target)
United Kingdom, Ireland	2001–2011	808 (29.7)	1,059 (38.9)	819 (30.1)	35 (1.3)	PCR (Ig/TCR)

Abbreviations: Ig/TCR, immunoglobulin heavy chain or T-cell receptor gene rearrangements; MRD, minimal residual disease; PCR, polymerase chain reaction.
[a]Eligible participants who died within 35 days of treatment start or who did not achieve remission were included in the overall survival and event-free survival analysis but were excluded from relapse and remission death analyses.

The UKALL2003 protocol was designed to test the feasibility of tailoring treatment intensity to relapse risk, based on MRD, in otherwise clinically standard- and intermediate-risk patients. The trial was essentially two RCTs in one study; MRD-HR patients were randomized to either standard or augmented treatment,[68] and in parallel, MRD-LR patients were randomized to either standard or reduced treatment.[69]

Clinically Standard- and Intermediate-Risk Patients Who Are MRD–Low-Risk

The 2013 article reports the results of reducing the delayed intensification phase of therapy in clinically standard- and intermediate-risk patients who have negligible or very low levels of MRD.[69] Of the 1,059 MRD-LR participants, 736 were eligible and 521 were randomized 1:1, balanced for age, sex, and white blood cell count. Table 14 outlines the study design for MRD-LR patients in the UKALL2003 protocol.

Table 14: Overview of Study Population and Groups in MRD-LR Patients in UKALL2003 Trial

Patients	No. of Study Sites	Outcomes	ALL Type, %	Age, years (MD, IQR)	% Male	Sample Size Reduced	Sample Size Standard
MRD-LR	45	1°: EFS[a] 2°: CIR, TR toxic effects, OS	B-cell 93 T-cell 7	4 (3–8)	57.6	260	261

Abbreviations: ALL, acute lymphoblastic leukemia; CIR, cumulative incidence of relapse; EFS, event-free survival; IQR, interquartile range; MD, median difference; MRD-LR, minimal residual disease–low risk; No., number; OS, overall survival; TR, treatment-related.
[a]EFS was defined as time to relapse, secondary tumour, or death in remission.
Source: Vora et al, 2013.[69]

Following the first few blocks of therapy (induction, consolidation, and interim maintenance) the reduced-treatment group received one delayed intensification followed by continuing therapy, whereas the standard-treatment group received two delayed intensifications separated by an interim maintenance phase, and then proceeded on to continuing therapy (full details are provided in Vora et al[69]).

The study groups were comparable after randomization in terms of clinical or biological characteristics, and the target sample size for adequate power was achieved and surpassed. Minimal residual disease was evaluated using real-time quantitative PCR for Ig/TCR rearrangements, in four laboratories adhering to a common quality assurance program, which reached a quantitative range of 10^{-4} (0.0001 or 1 in 10,000). Analyses were all intention-to-treat and actuarial percentages at 5 years.

Table 15: Results of Reduced Versus Standard Treatment in MRD-LR Patients in UKALL2003 Trial

Outcome	Reduced Group, % (95% CI)	Standard Group, % (95% CI)	Risk of Event in Standard Group, Odds Ratio (95% CI)[a]	P
5-year EFS[b]	94.4 (91.1–97.7)	95.5 (92.8–98.2)	1.00[b] (0.43–2.31)	.99
5-year OS[c]	97.9 (95.7–100)	98.5 (96.9–100)	0.67[c] (0.19–2.0)	.53
5-year relapse	5.6 (2.3–8.9)	2.4 (0.2–4.6)	0.55 (0.21–1.43)	.23

Abbreviations: ALL, acute lymphoblastic leukemia; CI, confidence interval; EFS, event-free survival; MRD-LR, minimal residual disease–low risk; OS, overall survival.
[a]Unadjusted values. Adjusted analysis was also conducted despite no evidence of baseline imbalance in the variables; randomization was balanced on: age (< 10 years vs. ≥ 10 years), sex (male vs. female), and white blood cell count at diagnosis (< 50 × 10^9/L vs. ≥ 50 × 10^9/L).
[b]Events included relapse, secondary tumour, or death.
[c]Event defined as any death.
Source: Vora et al, 2013.[69]

Minimal residual disease–low-risk patients were followed up for a median of 57 months from diagnosis (interquartile range 42–72) to assess outcomes. There were no significant differences between the standard-treatment and reduced-treatment groups in any of 5-year EFS, overall survival, and relapse risk (Table 15).[69] In addition to the unadjusted analysis in Table 15, the variables on which randomization was balanced (age, sex, and white blood cell count) were accounted for and did not yield materially different results. These findings were reported as stable following an additional 2 years of follow-up in the 2014 publication.[68]

Although not statistically significant, the proportion of the standard-treatment group that relapsed (2.4%) is nearly half that of the proportion in the reduced-treatment group (5.6%, P = .23).[69] Given the very low proportion of relapses in the trial overall, this numeric difference is unlikely to be worrisome. Experts advise that a treatment approach with two blocks of delayed intensification is no longer a common approach in North America. This review did not formally evaluate treatment-related toxicity; however, Vora et al[69] reported that there was no difference between the treatment groups in serious or life-threatening adverse events. This study provides evidence that MRD-directed treatment reduction in MRD-LR clinically standard- and intermediate-risk patients is feasible without compromise to patient outcomes. The GRADE assessment for this evidence is Moderate (Table A4).

Clinically Standard- and Intermediate-Risk Patients Who Are MRD–High-Risk

Results for MRD-HR patients were published later than those for MRD-LR patients because recruitment was extended (including an additional site) to enroll an adequate sample as prespecified in power calculations for this subpopulation.[68] Of the 808 MRD-HR patients, 533 were randomized to either standard- or augmented-treatment groups (Table 16). As with the MRD-LR randomization, groups were assigned on a 1:1 ratio, balanced for age, sex, and white blood cell count.

Table 16: Overview of Study Population and Groups in MRD-HR Patients in UKALL2003 Trial

Patients	No. of Study Sites	Outcomes	ALL Type(s)	Age, Years (MD, IQR)	% Male	Sample Size	
						Augmented	Standard
MRD-HR	46	1: EFS[a], OS	B-cell 86.1%	5	55.5	267	266
		2: CIR, TR toxic effects	T-cell 13.9%	(3–10)			

Abbreviations: ALL, acute lymphoblastic leukemia; CIR, cumulative incidence of relapse; EFS, event-free survival; IQR, interquartile range; MD, median; MRD-HR, minimal residual disease–high risk; No., number; OS, overall survival; TR, treatment-related.
[a]Events included relapse, secondary tumour, or death.
Source: Vora et al, 2014.[68]

The difference between the augmented and standard treatment was more doses of two and a differing dose of one other cytotoxic agent administered during postremission therapy (full details are provided in Vora et al[68]). Following randomization there were no significant differences in the clinical or biological characteristics of the study groups. As with all patients in the trial, MRD was evaluated in the same four laboratories using real-time quantitative PCR for Ig/TCR rearrangements, and statistical analyses were all intention-to-treat. The actuarial percentages at 5 years for EFS, overall survival, and relapse in the MRD-HR groups are summarized in Table 17. Planned subgroup analyses for these outcomes were also undertaken by known prognostic factors (sex, National Cancer Institute risk group, immunophenotype, MRD level on Day 29, and cytogenetic risk).

Table 17: Results of Augmented Versus Standard Treatment in MRD-HR Patients in UKALL2003 Trial

Outcome	Augmented Group,% (95% CI)	Standard Group,% (95% CI)	Risk of Event in Augmented Group, Odds Ratio[a] (95% CI)	P
5-year EFS[b]	89.6 (85.9–93.3)	82.2 (78.1–87.5)	0.61[b] (0.39–0.98)	.04
5-year OS[c]	92.9 (89.9–96.0)	88.9 (85.0–92.8)	0.67[c] (0.38–1.17)	.16
5-year relapse	7.5 (4.2–10.8)	14.2 (9.7–18.7)	0.55 (0.33–0.94)	.03

Abbreviations: ALL, acute lymphoblastic leukemia; CI, confidence interval; EFS, event-free survival; MRD-HR, minimal residual disease–high risk; OS, overall survival.
[a]Unadjusted odds ratio for augmented group; separate analysis accounting for the variables on which randomization was balanced: age (<10 years vs. ≥10 years), sex (male vs. female), and white blood cell count at diagnosis (< 50×10^9/L vs. ≥ 50×10^9/L) did not yield materially different results.
[b]Events included relapse, secondary tumour, or death.
[c]Event defined as any death.
Source: Vora et al, 2014.[68]

Outcomes were assessed in the MRD-HR patients upon completion of follow-up in October 2013, after a median of 70 months (interquartile range 52–91). Patients who received augmented treatment had a 39% superior EFS and 45% lower risk of relapse at 5 years compared with those who received standard therapy.[68] Among the relapses, the difference was driven by significantly fewer bone marrow relapses in the augmented group (4.6% vs. 10.5%, odds ratio = 0.42, P = .009), as there was no difference in non–bone marrow relapses between groups (P = .91).[68] No difference was observed in overall survival. None of the above results

differed across any subgroups when EFS, overall survival, and relapse were examined by prespecified prognostic factors.

The study observation window was 5 years, which could be too short to detect any change in overall survival resulting from the intervention. A difference in overall survival might not be detected because of relapsed patients who can be salvaged by HSCT, and because of relapsed patients who experience substantial morbidity and possibly subsequent mortality beyond the study period. Conducting a study with an observation period of decades or more is uncommon and challenging. The augmentation arm of this study included an escalated dose of intravenous methotrexate, which was shown to be superior to oral methotrexate for EFS and relapse risk reduction, which experts advise is in line with contemporary treatment protocols.

The authors of the UKALL2003 trial[68] identified another study that tested MRD-directed treatment augmentation and reduction (AIEOP-BFM-2000[43]) but noted that the results were not yet published, which was still the case to our knowledge at the time of writing. The ideal study design to determine the clinical utility of MRD as a prognostic factor would be an RCT in which all patients would be randomized to two arms: one arm with MRD evaluation and subsequent treatment selection (MRD + clinical risk stratification), and the other with clinical risk stratification and treatment selection only (marker-strategy RCT design; see Simon[71] for details). Although it suffers from some limitations, the UKALL2003 RCT is currently the best available evidence of patient-important outcomes using MRD-directed treatment. The GRADE quality of evidence for the body of evidence on relapse, EFS, and OS is Low to Moderate. Details of the GRADE for this body of evidence are in Appendix 2.

CONCLUSIONS

Despite heterogeneity in research methods (sample size, MRD cut-points, statistical analysis, and confounders included in the adjusted analyses), high or positive MRD levels in patients with ALL are a significant, independent prognostic factor for relapse when measured in each of the following scenarios:

- De novo ALL, evaluated at the end of induction (GRADE: Low)
- De novo ALL, evaluated at the end of consolidation (GRADE: Moderate)
- Relapsed ALL, evaluated after re-induction (GRADE: Moderate)
- Recipients of HSCTs, evaluated before transplantation (GRADE: Moderate)

In clinically standard- and intermediate-risk patients with ALL, MRD-directed treatment selection was beneficial in that:

- MRD–low-risk patients receiving MRD-directed treatment reduction experienced no compromise to EFS, overall survival, or relapse risk reduction compared with standard treatment (GRADE: Moderate)
- MRD–high-risk patients receiving MRD-directed treatment augmentation experienced a significant benefit in EFS and relapse risk reduction compared with standard treatment (GRADE: Moderate). The trial did not show a benefit in overall survival (GRADE: Low)

Limitations

The review does not reflect or evaluate every possible permutation and combination of timing and categorization of MRD evaluation in ALL, but provides a focused view of literature within a defined timeframe of publication. The date limitation of 1998 was selected on the basis of identification of landmark studies during scoping and consultation with local experts, in addition to decision-maker timeline and resource constraints. It is unknown how or if consideration of all studies since database inception would affect the results.

The most common outcome in the documented literature is EFS, with the primary determinant being relapse. A drawback of a precise focus on relapse itself is that the studies represent only a subset of the overall literature on MRD. A challenge, this is a trade-off between rigorous inclusion criteria that represent a subset of data, and more liberal criteria that could capture all studies on the topic. While focused, this review bolsters studies of EFS by connecting MRD as a prognostic factor discretely to relapse. The Foundation for the National Institutes of Health launched a 3-year study of MRD in fall 2014 that includes undertaking standardization of flow cytometry measurement of MRD and is expected to include a meta-analysis of existing MRD data.[28] The Foundation for the National Institutes of Health study, once available, could be complementary to this review.

A definitive characterization of the magnitude and precision of the independent prognostic value of MRD for relapse is yet to be attained. Clinical practice guidelines advising on the use of MRD do not currently exist. Our results do not allow for definitive identification of the optimal cut-point for positivity or of the role for MRD information in clinical practice (treatment selection). Most literature evaluated MRD with PCR, but flow cytometry will be used in Ontario. Experts advise that the results are comparable in terms of clinical decision-making, so this difference is unlikely to be an issue in practice.[30]

Evolution of MRD assays over time created inconsistency in the parameters, especially the cut-points for MRD positivity and negativity. Treatment also varied between studies, though similar drugs were generally administered (albeit at different doses, intensities, and frequencies). International collaborative study groups have pioneered advances in ALL treatment to improve patient outcomes. Treatment regimens used in the included studies might not be representative of contemporary protocols used locally.

Synthesis of heterogeneity in research methods proved challenging. For example, sample size, MRD cut-points, and confounders adjusted for in the analyses differed. The reporting of results of identified studies varied between studies; sometimes key data, such as standard error and even point estimates, were unavailable. Overall, there was a lack of adequate power among prognostic studies, likely resulting from the challenge of research on rare disease with few events (20%–30% of study participants relapsed; second relapses occurred in 43% of relapsed participants). Future studies could consider that, if patient-level data for these studies were made available, the heterogeneity could be overcome for statistical synthesis via individual patient data meta-analysis.

LIST OF ABBREVIATIONS

ALL	Acute lymphoblastic leukemia
AML	Acute myeloid leukemia
BFM	Berlin–Frankfurt–Münster study group
CNS	Central nervous system
DNA	Deoxyribonucleic acid
EFS	Event-free survival
GRADE	Grading of Recommendations Assessment, Development, and Evaluation
GVHD	Graft-versus-host disease
HSCT	Hematopoietic stem cell transplant
Ig/TCR	Immunoglobulin heavy chain or T-cell receptor [gene rearrangements]
MRD	Minimal residual disease
MRD-HR	MRD–high-risk
MRD-LR	MRD–low-risk
PCR	Polymerase chain reaction
Ph+	Philadelphia chromosome–positive
RCT	Randomized controlled trial

APPENDICES

Appendix 1: Literature Search Strategies

Database: EBM Reviews - Cochrane Database of Systematic Reviews <2005 to September 2014>, EBM Reviews - ACP Journal Club <1991 to October 2014>, EBM Reviews - Database of Abstracts of Reviews of Effects <3rd Quarter 2014>, EBM Reviews - Cochrane Central Register of Controlled Trials <September 2014>, EBM Reviews - Cochrane Methodology Register <3rd Quarter 2012>, EBM Reviews - Health Technology Assessment <3rd Quarter 2014>, EBM Reviews - NHS Economic Evaluation Database <3rd Quarter 2014>, Embase <1980 to 2014 Week 44>, Ovid MEDLINE(R) <1946 to October Week 4 2014>, Ovid MEDLINE(R) In-Process & Other Non-Indexed Citations <October 31, 2014>

Search Strategy:

#	Searches	Results
1	exp Precursor Cell Lymphoblastic Leukemia-Lymphoma/ use mesz,acp,cctr,coch,clcmr,dare,clhta,cleed or exp Leukemia, Myeloid, Acute/ use mesz,acp,cctr,coch,clcmr,dare,clhta,cleed or Leukemia, Myelomonocytic, Acute/ use mesz,acp,cctr,coch,clcmr,dare,clhta,cleed or Leukemia, Myelomonocytic, Juvenile/ use mesz,acp,cctr,coch,clcmr,dare,clhta,cleed	66892
2	exp acute leukemia/ use emez or exp childhood leukemia/ use emez	99567
3	Hematopoietic Stem Cell Transplantation/	54393
4	or/1-3	214761
5	Leukemia, B-Cell/ use mesz,acp,cctr,coch,clcmr,dare,clhta,cleed or Leukemia, T-Cell/ use mesz,acp,cctr,coch,clcmr,dare,clhta,cleed or Leukemia, Lymphoid/ use mesz,acp,cctr,coch,clcmr,dare,clhta,cleed or Leukemia, Myeloid/ use mesz,acp,cctr,coch,clcmr,dare,clhta,cleed	44904
6	B cell leukemia/ use emez or T cell leukemia/ use emez or lymphatic leukemia/ use emez or myeloid leukemia/ use emez	39720
7	or/5-6	84624
8	Acute Disease/	283628
9	7 and 8	9045
10	(((leuk?emi* or leuc?emi*) adj3 (lympho* or lymphat* or myelo* or granulocyt* or nonlympho* or promyelo* or megakaryoblast* or monocyt* or erythroblast* or B-cell or T-cell or B-ALL or T-ALL)) or ((childhood or precursor-B-cell) adj3 ALL) or ANLL or AML or (lymphoma adj lymphoblast*)).ti,ab.	315449
11	(acute or precursor or primary or relapse or recurren*).ti,ab.	5263365
12	10 and 11	182286
13	((hematopoietic adj5 stem cell transplant*) or HSCT).ti,ab.	34596
14	or/4,9,12-13	302309
15	Neoplasm, Residual/ use mesz,acp,cctr,coch,clcmr,dare,clhta,cleed	7446
16	minimal residual disease/ use emez	12508

17	(MRD or (residual adj3 (minimal or disease* or leuk?emi* or leuc?emi* or test*))).ti,ab.	35981
18	or/15-17	44118
19	14 and 18	9910
20	Case Reports/ or Comment.pt. or Editorial.pt. or Letter.pt. or Congresses.pt.	4319726
21	Case Report/ or Comment/ or Editorial/ or Letter/ or conference abstract.pt.	7577072
22	or/20-21	7657547
23	19 not 22	7334
24	limit 23 to english language [Limit not valid in CDSR,ACP Journal Club,DARE,CLCMR; records were retained]	6496
25	limit 24 to yr="1998 -Current" [Limit not valid in DARE; records were retained]	4988
26	remove duplicates from 25	2936

Appendix 2: Evidence Quality Assessment

Table A1: QUIPS[a] Risk of Bias Among Prognostic Studies Assessing the Independent Prognostic Value of MRD for Relapse

Author, Year	Study Participation	Study Attrition	Prognostic Factor Measurement	Outcome Measurement	Study Confounding	Statistical Analysis and Reporting
Meleshko et al, 2011[67]	Low	Low	Low	Low	Moderate[b]	Low
Zhou et al, 2007[70]	Low	Low	Moderate[c]	Low	Moderate[c]	Low
Laughton et al, 2005[66]	Low	Low	Low	Low	Moderate[d]	Low
Dworzak et al, 2002[64]	Low	Low	Moderate[e]	Low	Low	Low
Coustan-Smith et al, 2000[57]	Low	Low	Low	Low	Low	Moderate[f]
Van Dongen et al, 1998[58]	Low	Low	Low	Low	Moderate[d]	Low
Cavé et al, 1998[56]	Low	Low	Low	Low	Moderate[g]	Moderate[d]
Coustan-Smith et al, 2004[63]	Low	Low	Low	Low	Low	Low
Gandemer et al, 2014[65]	Low	Low	Low	Low	Low	Low
Balduzzi et al, 2014[61]	Low	Low	Low	Low	Low	Low
Bader et al, 2002[55]	Low	Low	Low	Low	Low	High[h]

Abbreviations: MRD, minimal residual disease; QUIPS, Quality In Prognostic Studies tool.

[a]Each consideration for the study is judged on information within the published article(s) to be at low, moderate, or high risk of bias. Details of QUIPS risk of bias assessment for studies are described in Hayden et al.[54]

[b]Observed effect of MRD on relapse could be distorted by another factor. Analysis did not adjust for either cytogenetics (owing to low occurrence of adverse cytogenetics) or treatment group, which differed by an additional randomization to one or two cytotoxic agent(s) during induction therapy.

[c]Observed effect of MRD on relapse could be distorted by another factor. Cut-points for MRD positivity used in the analysis were not defined a priori but were data-dependent, derived from recursive partitioning analysis, and cytogenetics were not adjusted for in the analysis.

[d]Observed effect of MRD on relapse could be distorted by another factor, as cytogenetics were not reported to be measured or accounted for in the analysis, so results could be spurious or biased related to analysis or reporting.

[e]A very low proportion (12%) of study participants experienced relapse that likely contributed to width of confidence interval, and 58.3% of study participants had complete data for the prognostic factor variable and were included in the analysis, which might not represent an adequate proportion of the study sample.

[f]No effect estimate, confidence interval, or standard error reported from prognostic analysis (*P* value only); analysis could have feasibly adjusted for multiple confounders simultaneously but did separate adjustment, and results could be spurious or biased because of analysis or reporting.

[g]Observed effect of MRD on relapse could be distorted by another factor. Data not shown for all analyses in publication. Some could not be acquired, and details were insufficient to include in meta-analysis.

[h]Statistical analysis and variables are poorly described, method of Cox model building is unclear, questionable number of covariates included in model considering sample size, and no effect estimate, confidence interval, or standard error reported from prognostic analysis (*P* value only), so results could be spurious or biased because of analysis or reporting.

Table A2: GRADE Evidence Profile for Independent Prognostic Value of MRD for Relapse

Number of Studies[a] (Design)	Risk of Bias	Inconsistency	Indirectness	Imprecision	Publication Bias	Upgrade Considerations	Quality
Relapse (End-induction)							
6 (observational)	Serious limitations (–1)[b]	No serious limitations[b]	No serious limitations	Serious limitations (–1)[c]	Undetected	None	⊕⊕ Low
Relapse (End-consolidation)							
2 (observational)	Serious limitations (–1)[d]	No serious limitations	No serious limitations	No serious limitations[e]	Undetected	None	⊕⊕⊕ Moderate
Relapse (After re-induction)							
1 (observational)	No serious limitations	No serious limitations	No serious limitations	Serious limitations (–1)[f]	Undetected	None	⊕⊕⊕ Moderate
Relapse (Pre-HSCT)							
2 (observational)	No serious limitations	Serious limitations (–1)[g]	No serious limitations	No serious limitations[h]	Undetected	None	⊕⊕⊕ Moderate

Abbreviations: GRADE, Grading of Recommendations Assessment, Development, and Evaluation; HSCT, hematopoietic stem cell transplant; MRD, minimal residual disease.

[a]A body of longitudinal cohort studies is considered to begin as high-quality evidence for prognosis according to GRADE guidance for assessment of evidence about prognosis detailed in Iorio et al.[53]

[b]All studies were prospective cohorts; median follow-up time was adequate in most studies; five studies[56,58,66,67,70] did not measure or adjust the analysis for all key confounders. See Table A1 for further detail on risk of bias assessment.

[c]Results across studies varied because of methodological heterogeneity, especially in thresholds for MRD positivity and adjusting for confounders precluding appropriate pooling. Relapse rate was low (range 12.0%–32.3%) and could decrease the statistical power, especially in smaller studies.[66,67] However, all effect estimates were in the same direction with overlapping confidence intervals (CIs) and were clinically significant.

[d]Both studies were prospective cohorts; median follow-up time was adequate but did not measure and/or adjust the analysis for all key confounders. See Table A1 for further detail on risk of bias assessment.

[e]No pooled estimate could be obtained because of methodological heterogeneity, and relapse rate per study was low (range 21.3%–25%), yielding questionable power. Thus confidence in the effect estimates is uncertain. However, clinical significance of the upper and lower bound of all CIs is retained despite wide CIs.

[f]Adequacy of follow-up is unknown, as the median was not reported. Sample size was very small (n = 35) with a relapse rate of 43%; thus, CIs are wide and there is uncertainty in the point estimate, though the clinical decision would be the same and clinical significance is retained at the upper and lower bound.

[g]Point estimates across studies varied by approximately 3-fold, and analyses were adjusted adequately for similar confounders; however, both effect estimates were in the same direction and were statistically and clinically significant, with overlapping CIs.

[h]Confidence intervals were very wide for both point estimates; adequate power to detect a clinically meaningful difference is unlikely because of low relapse rate (range 25.0%–26.8%), yielding uncertainty. However, the clinical decision would be the same and clinical significance is retained at the upper and lower bound.

Table A3: Risk of Bias Among Randomized Controlled Trials for Comparison of MRD-Directed Treatment Versus Standard Treatment

Author, Year	Allocation Concealment	Blinding	Complete Accounting of Patients and Outcome Events	Selective Reporting Bias	Other Limitations
Vora et al, 2014[68]	No limitations	No limitations[a]	No limitations	Limitations[b]	No limitations[c]
Vora et al, 2013[69]	No limitations	No limitations[a]	No limitations	Limitations[b]	No limitations[c]

Abbreviation: MRD, minimal residual disease.

[a]Patients and clinicians were not blinded to treatment because it was impossible, nor were data analysts; however, lack of blinding is unlikely to affect the outcomes of interest, which are survival-based, or events defined by morphologic diagnostic criteria (e.g., relapse).

[b]Study protocol pre-specified primary outcomes (event-free survival, overall survival) and secondary outcomes (remission rate, bone marrow relapse, non–bone marrow relapse, acute and late toxicity, days in hospital, and quality of life), for which insufficient data on subgroup analysis for primary outcomes are presented. One secondary outcome is not reported at all, and another is alluded to in the discussion of one article only.[68]

[c]Funders are reported to have had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

Table A4: GRADE Evidence Profile for Comparison of MRD-Directed Treatment Reduction Versus Standard Treatment in MRD–Low-Risk Patients

Number of Studies[a] (Design)	Risk of Bias	Inconsistency	Indirectness	Imprecision	Publication Bias	Upgrade Considerations	Quality
Event-Free Survival							
1 (RCT)	Serious limitations (–1)[b]	No serious limitations	No serious limitations	No serious limitations[c]	Undetected	None	⊕⊕⊕ Moderate
Overall Survival							
1 (RCT)	Serious limitations (–1)[b]	No serious limitations	No serious limitations	No serious limitations[d]	Undetected	None	⊕⊕⊕ Moderate
Relapse							
1 (RCT)	Serious limitations (–1)[b]	No serious limitations	No serious limitations	No serious limitations[c]	Undetected	None	⊕⊕⊕ Moderate

Abbreviations: GRADE, Grading of Recommendations Assessment, Development, and Evaluation; MRD, minimal residual disease; RCT, randomized controlled trial.

[a]Based on GRADE assessment for rating quality of evidence detailed in Guyatt et al.[52]

[b]See Table A3 for risk of bias assessment.

[c]Few events overall; however, target sample size was met and exceeded to achieve planned power to detect differences in primary outcomes, even after accounting for potential attrition. Authors state that ability to rule out a clinically significant 7% reduction in event-free survival was achieved by randomization.

[d]Few deaths during study observation period, adequately powered to rule out a 10% difference in overall survival (OS) between groups but unclear if time was sufficient to adequately evaluate OS.

Table A5: GRADE Evidence Profile for Comparison of MRD-Directed Treatment Augmentation Versus Standard Treatment in MRD–High-Risk Patients

Number of Studies[a] (Design)	Risk of Bias	Inconsistency	Indirectness	Imprecision	Publication Bias	Upgrade Considerations	Quality
Event-Free Survival							
1 (RCT)	Serious limitations (–1)[b]	No serious limitations	No serious limitations	No serious limitations[c]	Undetected	None	⊕⊕⊕ Moderate
Overall Survival							
1 (RCT)	Serious limitations (–1)[b]	No serious limitations	No serious limitations	Serious limitations (–1)[d]	Undetected	None	⊕⊕ Low
Relapse							
1 (RCT)	Serious limitations (–1)[b]	No serious limitations	No serious limitations	No serious limitations[c]	Undetected	None	⊕⊕⊕ Moderate

Abbreviations: GRADE, Grading of Recommendations Assessment, Development, and Evaluation; MRD, minimal residual disease; RCT, randomized controlled trial.

[a]Based on GRADE assessment for rating quality of evidence detailed in Guyatt et al.[52]

[b]See Table A3 for risk of bias assessment.

[c]Few events and technically underpowered to detect effect sizes found; however, target sample size for > 80% power to detect a 10% difference between groups after accounting for potential attrition was met, and between-group differences are clinically important and statistically significant.

[d]Few deaths during study observation period yielded imprecision around the estimate and lack of power to adequately evaluate a clinically meaningful difference between treatment groups in overall survival.

REFERENCES

(1)   Leukemia & Lymphoma Society of Canada. Leukemia [Internet]. 2013 [updated 10 July 2013]. Available from: http://www.llscanada.org/#/diseaseinformation/leukemia/

(2)   Canadian Cancer Society's Advisory Committee on Cancer Statistics. Canadian Cancer Statistics 2014 [Internet]. Toronto, ON: Canadian Cancer Society; 2014 May [cited 2014 Nov 21]. Available from: http://www.cancer.ca/~/media/cancer.ca/CW/cancer%20information/cancer%20101/Canadian%20cancer%20statistics/Canadian-Cancer-Statistics-2014-EN.pdf

(3)   National Cancer Institute. PDQ® childhood acute lymphoblastic leukemia treatment [Internet]. Bethesda (MD): The Institute; 2014 Nov [cited 2014 Nov 21]. Available from: http://cancer.gov/cancertopics/pdq/treatment/childALL/Patient

(4)   National Cancer Institute. PDQ Childhood Acute Myeloid Leukemia/Other Myeloid Malignancies Treatment [Internet]. Bethesda, MD: National Cancer Institute; 2014 Nov [cited 2014 Nov 21]. Available from: http://cancer.gov/cancertopics/pdq/treatment/childAML/Patient

(5)   Macleod RA, Kaufmann M, Drexler HG. Cytogenetic analysis of cancer cell lines. Methods Mol Biol. 2011;731:57-78.

(6)   Pui CH, Evans WE. Acute lymphoblastic leukemia. N Engl J Med. 1998;339(9):605-15.

(7)   Cazzaniga G, Biondi A. Molecular monitoring of childhood acute lymphoblastic leukemia using antigen receptor gene rearrangements and quantitative polymerase chain reaction technology. Haematologica. 2005;90(3):382-90.

(8)   Pui CH, Relling MV, Downing JR. Acute lymphoblastic leukemia. N Engl J Med. 2004;350(15):1535-48.

(9)   Locatelli F, Schrappe M, Bernardo ME, Rutella S. How I treat relapsed childhood acute lymphoblastic leukemia. Blood. 2012;120(14):2807-16.

(10)   Moppett J, Burke GA, Steward CG, Oakhill A, Goulden NJ. The clinical relevance of detection of minimal residual disease in childhood acute lymphoblastic leukaemia. J Clin Pathol. 2003;56(4):249-53.

(11)   Onciu M. Acute lymphoblastic leukemia. Hematol Oncol Clin North Am. 2009;23(4):655-74.

(12)   zur Stadt U, Harms DO, Schluter S, Schrappe M, Goebel U, Spaar H, et al. MRD at the end of induction therapy in childhood acute lymphoblastic leukemia: outcome prediction strongly depends on the therapeutic regimen. Leukemia. 2001;15(2):283-5.

(13)   Pui CH, Howard SC. Current management and challenges of malignant disease in the CNS in paediatric leukaemia. Lancet Oncol. 2008;9(3):257-68.

(14)   Raetz EA, Borowitz MJ, Devidas M, Linda SB, Hunger SP, Winick NJ, et al. Reinduction platform for children with first marrow relapse of acute lymphoblastic leukemia: a Children's Oncology Group Study [corrected]. J Clin Oncol. 2008;26(24):3971-8.

(15)   Henderson MJ, Choi S, Beesley AH, Sutton R, Venn NC, Marshall GM, et al. Mechanism of relapse in pediatric acute lymphoblastic leukemia. Cell Cycle. 2008;7(10):1315-20.

(16)   Mullighan CG, Phillips LA, Su X, Ma J, Miller CB, Shurtleff SA, et al. Genomic analysis of the clonal origins of relapsed acute lymphoblastic leukemia. Science. 2008;322(5906):1377-80.

(17)   Satwani P, Sather H, Ozkaynak F, Heerema NA, Schultz KR, Sanders J, et al. Allogeneic bone marrow transplantation in first remission for children with ultra-high-risk features of acute lymphoblastic leukemia: a Children's Oncology Group study report. Biol Blood Marrow Transplant. 2007;13(2):218-27.

(18)   Wingard JR, Majhail NS, Brazauskas R, Wang Z, Sobocinski KA, Jacobsohn D, et al. Long-term survival and late deaths after allogeneic hematopoietic cell transplantation. J Clin Oncol. 2011;29(16):2230-9.

(19)   Kotecha RS, Gottardo NG, Kees UR, Cole CH. The evolution of clinical trials for infant acute lymphoblastic leukemia. Blood Cancer J. 2014;4:e200.

(20)   Pediatric Oncology Group of Ontario MRD Provincial Working Group. Minimal residual disease (MRD) and childhood leukemia: standard of care position paper. Toronto (ON): Pediatric Oncology Group of Ontario; 2014.

(21)   National Cancer Institute. Blood-Forming Stem Cell Transplants (Fact Sheet) [Internet]. Rockville, MD: National Cancer Institute; 2013 [updated 12 August 2013]. Available from: http://www.cancer.gov/cancertopics/treatment/types/stem-cell-transplant/stem-cell-fact-sheet.

(22)   Barton-Burke M, Dwinell DM, Kafkas L, Lavalley C, Sands H, Proctor C, et al. Graft-versus-host disease: a complex long-term side effect of hematopoietic stem cell transplant. Oncology (Williston Park). 2008;22(11 Suppl Nurse Ed):31-45.

(23)   Friedmann AM, Weinstein HJ. The role of prognostic features in the treatment of childhood acute lymphoblastic leukemia. Oncologist. 2000;5(4):321-8.

(24)   Pui CH, Pei D, Campana D, Bowman WP, Sandlund JT, Kaste SC, et al. Improved prognosis for older adolescents with acute lymphoblastic leukemia. J Clin Oncol. 2011;29(4):386-91.

(25)   Campana D. Minimal residual disease monitoring in childhood acute lymphoblastic leukemia. Curr Opin Hematol. 2012;19(4):313-8.

(26)   Campana D. Minimal residual disease in acute lymphoblastic leukemia. Hematology Am Soc Hematol Educ Program. 2010;2010:7-12.

(27)   Campana D. Role of minimal residual disease monitoring in adult and pediatric acute lymphoblastic leukemia. Hematol Oncol Clin North Am. 2009;23(5):1083-98, vii.

(28)   Foundation for the National Institutes of Health. MRD Press Release [Internet]. Bethesda, MD: Foundation for the National Institutes of Health; 2014 Oct [cited 2015 May 7]. Available from: http://www.fnih.org/sites/all/files/documents/FNIH_MRD_final_1008.pdf

(29)   Bruggemann M, Schrauder A, Raff T, Pfeifer H, Dworzak M, Ottmann OG, et al. Standardized MRD quantification in European ALL trials: proceedings of the Second International Symposium on MRD assessment in Kiel, Germany, 18-20 September 2008. Leukemia. 2010;24(3):521-35.

(30)   Gaipa G, Basso G, Biondi A, Campana D. Detection of minimal residual disease in pediatric acute lymphoblastic leukemia. Cytometry B Clin Cytom. 2013;84(6):359-69.

(31)   Brisco MJ, Condon J, Hughes E, Neoh SH, Sykes PJ, Seshadri R, et al. Outcome prediction in childhood acute lymphoblastic leukaemia by molecular quantification of residual disease at the end of induction. Lancet. 1994;343(8891):196-200.

(32)   Riley RD, Sauerbrei W, Altman DG. Prognostic markers in cancer: the evolution of evidence from single studies to meta-analysis, and beyond. Br J Cancer. 2009;100(8):1219-29.

(33)   Tezak Z, Kondratovich M, Mansfield E. US FDA and personalized medicine: in vitro diagnostic regulatory perspective. Per Med. 2010;7(5):517-30.

(34)   Italiano A. Prognostic or predictive? It's time to get back to definitions! J Clin Oncol. 2011;29(35):4718-9.

(35)   Oldenhuis CN, Oosting SF, Gietema JA, de Vries EG. Prognostic versus predictive value of biomarkers in oncology. Eur J Cancer. 2008;44(7):946-53.

(36)   Hingorani AD, Windt DA, Riley RD, Abrams K, Moons KG, Steyerberg EW, et al. Prognosis research strategy (PROGRESS) 4: stratified medicine research. BMJ. 2013;346:e5793.

(37)   Altman DG. Systematic reviews of evaluations of prognostic variables. BMJ. 2001;323(7306):224-8.

(38)   Riley RD, Abrams KR, Sutton AJ, Lambert PC, Jones DR, Heney D, et al. Reporting of prognostic markers: current problems and development of guidelines for evidence-based practice in the future. Br J Cancer. 2003;88(8):1191-8.

(39)   Kattan MW. Evaluating a new marker's predictive contribution. Clin Cancer Res. 2004;10(3):822-4.

(40)   Peduzzi P, Concato J, Feinstein AR, Holford TR. Importance of events per independent variable in proportional hazards regression analysis. II. Accuracy and precision of regression estimates. J Clin Epidemiol. 1995;48(12):1503-10.

(41)   Simon R, Altman DG. Statistical aspects of prognostic factor studies in oncology. Br J Cancer. 1994;69(6):979-85.

(42)   Borowitz MJ, Devidas M, Hunger SP, Bowman WP, Carroll AJ, Carroll WL, et al. Clinical significance of minimal residual disease in childhood acute lymphoblastic leukemia and its relationship to other prognostic factors: a Children's Oncology Group study. Blood. 2008;111(12):5477-85.

(43)   Conter V, Bartram CR, Valsecchi MG, Schrauder A, Panzer-Grumayer R, Moricke A, et al. Molecular response to treatment redefines all prognostic factors in children and adolescents with B-cell precursor acute lymphoblastic leukemia: results in 3184 patients of the AIEOP-BFM ALL 2000 study. Blood. 2010;115(16):3206-14.

(44)   Schrappe M, Valsecchi MG, Bartram CR, Schrauder A, Panzer-Grumayer R, Moricke A, et al. Late MRD response determines relapse risk overall and in subsets of childhood T-cell ALL: results of the AIEOP-BFM-ALL 2000 study. Blood. 2011;118(8):2077-84.

(45)   Van der Velden VH, Corral L, Valsecchi MG, Jansen MW, De Lorenzo P, Cazzaniga G, et al. Prognostic significance of minimal residual disease in infants with acute lymphoblastic leukemia treated within the Interfant-99 protocol. Leukemia. 2009;23(6):1073-9.

(46)   Eckert C, von Stackelberg A, Seeger K, Groeneveld TW, Peters C, Klingebiel T, et al. Minimal residual disease after induction is the strongest predictor of prognosis in intermediate risk relapsed acute lymphoblastic leukaemia—long-term results of trial ALL-REZ BFM P95/96. Eur J Cancer. 2013;49(6):1346-55.

(47)   Eckert C, Henze G, Seeger K, Hagedorn N, Mann G, Panzer-Grumayer R, et al. Use of allogeneic hematopoietic stem-cell transplantation based on minimal residual disease response improves outcomes for children with relapsed acute lymphoblastic leukemia in the intermediate-risk group. J Clin Oncol. 2013;31(21):2736-42.

(48)   Sutton R, Shaw PJ, Venn NC, Law T, Dissanayake A, Kilo T, et al. Persistent MRD before and after allogeneic BMT predicts relapse in children with acute lymphoblastic leukaemia. Br J Haematol. 2015;168(3):395-404.

(49)   Fleming TR, Rothmann MD, Lu HL. Issues in using progression-free survival when evaluating oncology products. J Clin Oncol. 2009;27(17):2874-80.

(50)   McCain Jr. J. The ongoing evolution of endpoints in oncology. Managed Care (Suppl 1). 2010;19(5):11.

(51)   Smith BD, Karp JE. What are the endpoints of therapy for acute leukemias? Old definitions and new challenges. Clin Lymphoma Myeloma. 2009;9 Suppl 3:S296-301.

(52)   Guyatt GH, Oxman AD, Schunemann HJ, Tugwell P, Knottnerus A. GRADE guidelines: a new series of articles in the Journal of Clinical Epidemiology. J Clin Epidemiol. 2011;64(4):380-2.

(53)   Iorio A, Spencer FA, Falavigna M, Alba C, Lang E, Burnand B, et al. Use of GRADE for assessment of evidence about prognosis: rating confidence in estimates of event rates in broad categories of patients. BMJ. 2015;350:h870.

(54) Hayden JA, Cote P, Bombardier C. Evaluation of the quality of prognosis studies in systematic reviews. Ann Intern Med. 2006;144(6):427-37.

(55) Bader P, Hancock J, Kreyenberg H, Goulden NJ, Niethammer D, Oakhill A, et al. Minimal residual disease (MRD) status prior to allogeneic stem cell transplantation is a powerful predictor for post-transplant outcome in children with ALL. Leukemia. 2002;16(9):1668-72.

(56) Cavé H, Van Der Werff Ten Bosch J, Suciu S, Guidal C, Waterkeyn C, Otten J, et al. Clinical significance of minimal residual disease in childhood acute lymphoblastic leukemia. N Engl J Med. 1998;339(9):591-8.

(57) Coustan-Smith E, Sancho J, Hancock ML, Boyett JM, Behm FG, Raimondi SC, et al. Clinical importance of minimal residual disease in childhood acute lymphoblastic leukemia. Blood. 2000;96(8):2691-6.

(58) van Dongen JJ, Seriu T, Panzer-Grumayer ER, Biondi A, Pongers-Willemse MJ, Corral L, et al. Prognostic value of minimal residual disease in acute lymphoblastic leukaemia in childhood. Lancet. 1998;352(9142):1731-8.

(59) Knechtli CJ, Goulden NJ, Hancock JP, Grandage VL, Harris EL, Garland RJ, et al. Minimal residual disease status before allogeneic bone marrow transplantation is an important determinant of successful outcome for children and adolescents with acute lymphoblastic leukemia. Blood. 1998;92(11):4072-9.

(60) Sramkova L, Muzikova K, Fronkova E, Krejci O, Sedlacek P, Formankova R, et al. Detectable minimal residual disease before allogeneic hematopoietic stem cell transplantation predicts extremely poor prognosis in children with acute lymphoblastic leukemia. Pediatr Blood Cancer. 2007;48(1):93-100.

(61) Balduzzi A, Di Maio L, Silvestri D, Songia S, Bonanomi S, Rovelli A, et al. Minimal residual disease before and after transplantation for childhood acute lymphoblastic leukaemia: is there any room for intervention? Br J Haematol. 2014;164(3):396-408.

(62) Coustan-Smith E, Behm FG, Sanchez J, Boyett JM, Hancock ML, Raimondi SC, et al. Immunological detection of minimal residual disease in children with acute lymphoblastic leukaemia. Lancet. 1998;351(9102):550-4.

(63) Coustan-Smith E, Gajjar A, Hijiya N, Razzouk BI, Ribeiro RC, Rivera GK, et al. Clinical significance of minimal residual disease in childhood acute lymphoblastic leukemia after first relapse. Leukemia. 2004;18(3):499-504.

(64) Dworzak MN, Froschl G, Printz D, Mann G, Potschger U, Muhlegger N, et al. Prognostic significance and modalities of flow cytometric minimal residual disease detection in childhood acute lymphoblastic leukemia. Blood. 2002;99(6):1952-8.

(65) Gandemer V, Pochon C, Oger E, Dalle JH, Michel G, Schmitt C, et al. Clinical value of pre-transplant minimal residual disease in childhood lymphoblastic leukaemia: the results of the French minimal residual disease-guided protocol. Br J Haematol. 2014;165(3):392-401.

(66) Laughton SJ, Ashton LJ, Kwan E, Norris MD, Haber M, Marshall GM. Early responses to chemotherapy of normal and malignant hematologic cells are prognostic in children with acute lymphoblastic leukemia. J Clin Oncol. 2005;23(10):2264-71.

(67) Meleshko AN, Savva NN, Fedasenka UU, Romancova AS, Krasko OV, Eckert C, et al. Prognostic value of MRD-dynamics in childhood acute lymphoblastic leukemia treated according to the MB-2002/2008 protocols. Leuk Res. 2011;35(10):1312-20.

(68) Vora A, Goulden N, Mitchell C, Hancock J, Hough R, Rowntree C, et al. Augmented post-remission therapy for a minimal residual disease-defined high-risk subgroup of children and young people with clinical standard-risk and intermediate-risk acute lymphoblastic leukaemia (UKALL 2003): a randomised controlled trial. Lancet Oncol. 2014;15(8):809-18.

(69)   Vora A, Goulden N, Wade R, Mitchell C, Hancock J, Hough R, et al. Treatment reduction
        for children and young adults with low-risk acute lymphoblastic leukaemia defined by
        minimal residual disease (UKALL 2003): a randomised controlled trial. Lancet Oncol.
        2013;14(3):199-209.
(70)   Zhou J, Goldwasser MA, Li A, Dahlberg SE, Neuberg D, Wang H, et al. Quantitative
        analysis of minimal residual disease predicts relapse in children with B-lineage acute
        lymphoblastic leukemia in DFCI ALL Consortium Protocol 95-01. Blood.
        2007;110(5):1607-11.
(71)   Simon R. Clinical trial designs for evaluating the medical utility of prognostic and
        predictive biomarkers in oncology. Per Med. 2010;7(1):33-47.

About Health Quality Ontario

Health Quality Ontario is the provincial advisor on the quality of health care. We are motivated by a single-minded purpose: **Better health for all Ontarians.**

Who We Are.

We are a scientifically rigorous group with diverse areas of expertise. We strive for complete objectivity, and look at things from a vantage point that allows us to see the forest and the trees. We work in partnership with health care providers and organizations across the system, and engage with patients themselves, to help initiate substantial and sustainable change to the province's complex health system.

What We Do.

We define the meaning of quality as it pertains to health care, and provide strategic advice so all the parts of the system can improve. We also analyze virtually all aspects of Ontario's health care. This includes looking at the overall health of Ontarians, how well different areas of the system are working together, and most importantly, patient experience. We then produce comprehensive, objective reports based on data, facts and the voice of patients, caregivers and those who work each day in the health system. As well, we make recommendations on how to improve care using the best evidence. Finally, we support large scale quality improvements by working with our partners to facilitate ways for health care providers to learn from each other and share innovative approaches.

Why It Matters.

We recognize that, as a system, we have much to be proud of, but also that it often falls short of being the best it can be. Plus certain vulnerable segments of the population are not receiving acceptable levels of attention. Our intent at Health Quality Ontario is to continuously improve the quality of health care in this province regardless of who you are or where you live. We are driven by the desire to make the system better, and by the inarguable fact that better has no limit.

Permission Requests: All inquiries regarding permission to reproduce any content in Health Quality Ontario reports should be directed to EvidenceInfo@hqontario.ca.

About Health Quality Ontario

About the Ontario Health Technology Advisory Committee (OHTAC)

How to Obtain OHTAC Recommendation Reports

Health Quality Ontario
130 Bloor Street West, 10th Floor
Toronto, Ontario
M5S 1N5
Tel: 416-323-6868
Toll Free: 1-866-623-6868
Fax: 416-323-9261
Email: EvidenceInfo@hqontario.ca
www.hqontario.ca