Early Response And Minimal Residual Disease Testing In Childhood ALL: Methodologies And Clinical Application

Stephen P. Hunger, Andrea Biondi

The clinical utility of early response measures
Approximately 80% of children with acute lymphoblastic leukemia (ALL) are cured with contemporary risk-directed treatment, with therapies of different intensities administered to patient groups with differential risks of relapse. Many different clinical and biological features are used to identify different risk groups. Early response to therapy, defined as the initial degree and rate of disease regression prior to end induction, is one of the most powerful. Simple enumeration of blasts remaining in peripheral blood (PB) or bone marrow (BM) at defined times during induction therapy is highly predictive of treatment outcome among large groups of patients with ALL.\(^1\)

The Berlin-Frankfurt-Munster (BFM) group showed in ALL-BFM 83 that the number of blasts remaining in the peripheral blood following seven days treatment with prednisone and a single dose of intrathecal methotrexate separates patients into two subgroups with divergent outcomes.\(^2\) Prednisone good responders (PGR) consist of the approximately 90% of patients that have < 1000 blasts/ml remaining in the PB and have an excellent outcome. In contrast, prednisone poor responders (PPR) that have = 1000 blasts/ml PB have an extremely poor outcome. Prednisone response has been a cornerstone of risk stratification in BFM and most other European ALL cooperative group clinical trials for the past several decades.

In parallel, Children’s Cancer Group (CCG) studies conducted in the US and Canada used BM morphology to assess early response and showed significantly worse outcomes for patients with an M3 marrow (= 25% blasts) versus an M1 (<5% blasts) or M2 (5-25% blasts) marrow at day 7 or 14 of induction therapy.\(^3,4\) The CCG also showed that early response could be used to target new therapeutic strategies to subgroups of patients at high risk of relapse. CCG 1882 randomized high risk ALL patients (age = 10 years and/or initial white blood cell count = 50,000/microliter) that had a day 7 M3 marrow to receive either augmented or standard therapy and found that augmented therapy led to major improvements in treatment outcome.\(^5\) Once a therapy is proven safe and effective in a high-risk subgroup, it can then be tested in patients with a lower risk of relapse. Pursuing this strategy, the CCG tested augmented therapy in high risk ALL patients with a day 7 M1 or M2 marrow in CCG 1961 and found it to be significantly more effective than standard therapy.\(^6\) Similar augmented therapy is now being tested in a large subset of standard risk ALL patients in the Children’s Oncology Group (COG) AALL0331 trial.

While poor early response as assessed by PB or BM morphology following 1-2 weeks of therapy can identify patient subgroups at significantly increased risk of relapse, this approach has significant limitations. Many patients with good responses relapse, while many of those with poor responses are cured. Indeed, about 70% of relapses in BFM trials occur in PGR patients.\(^7\) Over the past 15-20 years, more sensitive molecular biology and flow cytometric technologies have been developed to assess early response by measuring subclinical levels of minimal residual disease (MRD). To date, modest sized, mostly retrospective, studies have demonstrated that BM MRD measurement in the first 1-3 months of therapy may separate patients into very good and very poor risk subsets better than conventional measures of early response. Studies now in progress will help to define whether or not MRD assessment leads to major improvements in risk stratification and treatment allocation in large clinical trials.
Clinical features (age, white blood cell count), genetic properties of the leukemia cells (presence/absence of specific translocations or chromosome gains/losses) and the host (polymorphisms in key genes) also have major impact on outcome. A major challenge is to define how early response should be integrated with other risk factors to develop the most robust treatment assignment algorithm. The fact that MRD is predictive of outcome in multivariate analyses establishes that it is not simply a surrogate for other prognostic factors.

Methods for detection of MRD in ALL

Two different technological approaches have been used to measure MRD in ALL: molecular methods that utilize the polymerase chain reaction (PCR) to amplify and quantitate tumor-specific or tumor-associated genetic markers, and quantification of residual leukemia cells with tumor-specific or tumor-associated phenotypes by flow cytometry. Each has relative advantages and disadvantages. Molecular methods generally have greater sensitivity, while flow-based methods are faster and less expensive. Studies have shown consistently that end induction MRD burden correlates with outcome in childhood ALL; the higher the MRD level, the worse the outcome.(8-12)

PCR can be used to measure MRD in several different ways. The immunoglobulin (Ig) heavy and light chain and T cell receptor (TCR) genes, collectively termed antigen receptor (AgRec) genes, undergo ordered recombination during B- and T-cell development thereby creating patient-specific markers of the malignant clone. Using AgRec targets that include the Ig heavy chain (IgH), TCR-delta, TCR-gamma and Ig-kappa genes at least one AgRec PCR target can be identified in 98% and at least two AgRec PCR targets in 95% of childhood B-precursor ALLs.(13) In most cases, MRD at levels of at least $10^{-4}$ (1 cell in 10,000), and often $10^{-5}$, can be detected using AgRec PCR. A second approach for molecular MRD detection is amplification of fusion RNA transcripts generated by leukemia-specific chromosome translocations via reverse-transcriptase PCR (RT-PCR). About 30-35% of childhood ALLs contain $BCR-ABL$, $MLL-AF4$, $E2A-PBX1$, or $TEL-AML1$ fusion transcripts. Because these transcripts are leukemia-specific, they clearly distinguish malignant from normal cells. Real-time quantitative PCR (RQ-PCR) methods can be used to detect and quantify these transcripts with high sensitivity and reliability.(14) However, because they are only present in subsets of patients, it is difficult to use them in large scale clinical trials, unless the trials are focused on distinct molecular subsets of ALL that receive unique therapies, such as Philadelphia chromosome-positive ALL.

Detection of MRD via flow cytometry is based on the principle that leukemia cells express patterns of antigens that differ from those of normal cells.(15) These differences are due to leukemia cells expressing novel combinations of antigens that are not encountered during normal lymphoid differentiation or expressing normal antigens in intensity patterns not observed during normal maturation.(16) Normal precursor B-cells express a precise and reproducible sequence of antigens during maturation. Using a modest number of 3- or 4-color combinations, leukemia cells can be found to occupy regions of so-called “empty space” on bivariate displays, where normal B cell precursors are not observed.(17) Practically speaking, flow cytometry can detect leukemia cells at a sensitivity in the range of $10^{-4}$ at end induction, and cooperative group experience finds that over 95% of B-precursor ALLs have informative phenotypes.(17-19)

Use of MRD measures in cooperative group clinical trials

North American and European cooperative groups have used very different strategies to integrate MRD measures into risk stratification algorithms in contemporary ALL clinical trials. In Europe, the joint Italian AEIOP, German and Austrian BFM group ALL-BFM-2000 trial uses MRD determined by AgRec PCR at end induction and end consolidation therapy to adjust treatment intensity.(20) A substantial amount of work was required to standardize MRD testing at several different laboratories in the three countries, as pioneered by the BIOMED initiative.(21) While this study is still accruing patients, it has established that it is feasible to integrate molecular MRD into large-scale cooperative group clinical trials. Over 80% of patients have been risk-stratified on the basis of
MRD using AgRec targets with a sensitivity of at least 10^{-4}. Several different therapy questions are being tested in this trial. Standard risk patients are defined as those with a PGR, no t(4;11), t(9;22) or hypodiploidy, and negative MRD (with a sensitivity of at least 10^{-4}) at end induction and end consolidation. Patients in this group, which includes about 30% of childhood ALL patients, are randomized to receive standard BFM therapy versus a reduced treatment regimen that decreases anthracycline dose in phase IIa (delayed intensification), with a goal of maintaining EFS at levels seen in the BFM 95 trial. At the other end of the spectrum, poor risk patients include those, regardless of initial presenting characteristics, that have MRD >1% at end consolidation. Those patients receive either a very intensive chemotherapy regimen or allogeneic stem cell transplant in an attempt to improve what is predicted to be a less than 50% cure rate.

The North American cooperative groups have also used MRD for risk stratification, but with a very different approach. The Paediatric Oncology Group (POG) established in the recently completed to accrual POG 9900 trial that MRD could be determined at end induction via flow cytometry in over 95% of patients. The new generation of COG ALL clinical trials use flow cytometry to assess MRD at end induction, with a primary goal of identifying additional high risk patients for treatment intensification. Patients are assigned to risk groups primarily on the basis of age and initial white blood cell count, genetic features of the leukemia cells, and conventional day 8/15 measures of early response. However, those with end induction MRD burdens of = 0.1% are non-randomly assigned to receive more intensive therapy. In addition, a separate low risk group is identified that includes NCI standard risk ALL patients with favorable cytogenetics (TEL-AML1 fusion or simultaneous trisomies of chromosomes 4, 10 and 17), no CNS or testicular disease, and an excellent early response defined by <5% marrow blasts by day 15 and MRD < 0.1% at the end of induction. These children will be randomized to receive a "minimal" standard therapy with or without four additional doses of PEG Asparaginase in the first 12 weeks following induction, with a goal of increasing the 6-year EFS from 90% to 94%.

Summary and perspective
Thus, studies currently in progress will define how useful it is to integrate MRD into ALL clinical trials and will assess the relative advantages and disadvantages of different technical approaches to MRD detection. Major contemporary challenges in the treatment of children with ALL are to identify those patients with an extremely high likelihood of being cured with currently available therapies so that adverse effects of further treatment intensification can be avoided, to identify patients unlikely to be cured with currently available therapies so that they can receive more intensive or novel therapies, and a need to develop better tools with which to measure treatment response to novel molecularly targeted therapies. Accurate measures of early treatment response will play a critical role in addressing each of these challenges.

However, we must also emphasize that one of the most noteworthy developments in contemporary paediatric oncology is organized efforts of resource rich countries to work with resource poor countries to develop systems to provide effective treatment for malignancies such as ALL. We must not lose sight of the fact that simple early response measures that can be performed anywhere, prednisone response and/or assessment of early marrow response, can provide critical information that can be used effectively to stratify therapy intensity based on risk of relapse. Risk stratified therapy may have particularly large benefits in this setting. Toxic death rates from intensive therapy are far higher than observed in resource rich countries, necessitating that these therapies be used cautiously and in a targeted manner. We must not establish a paradigm that suggests that only those centers that can measure MRD can provide effective contemporary therapy.

References


