CLINICAL SIGNIFICANCE OF MINIMAL RESIDUAL DISEASE IN CHILDHOOD ACUTE LYMPHOBLASTIC LEUKEMIA AND ITS RELATIONSHIP TO OTHER PROGNOSTIC FACTORS. A CHILDREN’S ONCOLOGY GROUP STUDY.

Michael J Borowitz\textsuperscript{1}, Meenakshi Devidas\textsuperscript{2}, Stephen P Hunger\textsuperscript{3}, W.Paul Bowman\textsuperscript{4}, Andrew J Carroll\textsuperscript{5}, William L Carroll\textsuperscript{6}, Stephen Linda\textsuperscript{2}, Paul L Martin\textsuperscript{7}, D. Jeanette Pullen\textsuperscript{8}, David Viswanatha \textsuperscript{9}, Cheryl L Willman\textsuperscript{10}, Naomi Winick\textsuperscript{11} and Bruce M Camitta\textsuperscript{12}.

\textsuperscript{1}Department of Pathology, Johns Hopkins Medical Institutions, Baltimore MD
\textsuperscript{2}Children’s Oncology Group and the College of Medicine, University of Florida, Gainesville FL
\textsuperscript{3}University of Colorado Denver and The Children's Hospital, Aurora CO
\textsuperscript{4}Cook Children’s Hospital, Fort Worth TX
\textsuperscript{5}University of Alabama at Birmingham, Birmingham, AL
\textsuperscript{6}New York University School of Medicine, New York NY
\textsuperscript{7}Duke University School of Medicine, Durham NC
\textsuperscript{8}University of Mississippi School of Medicine, Jackson MS
\textsuperscript{9}Mayo Clinic, Rochester, MN
\textsuperscript{10}University of New Mexico School of Medicine, Albuquerque NM
\textsuperscript{11}University of Texas Southwestern School of Medicine, Dallas, TX
\textsuperscript{12}Midwest Children’s Cancer Center, Medical College of Wisconsin, Milwaukee WS

And the Children’s Oncology Group, Arcadia, CA.

Running head: MRD in Childhood ALL

Supported by Grants R01 CA86011 and U10 CA98543 from NCI
Presented in part at the 2006 American Society of Hematology Meeting, Orlando FL
Minimal residual disease (MRD) is an important predictor of relapse in acute lymphoblastic leukemia (ALL) but its relationship to other prognostic variables has not been fully assessed. The Children’s Oncology Group studied the prognostic impact of MRD measured by flow cytometry in the peripheral blood at day 8, and in end induction (day 29) and end consolidation marrows in 2143 children with precursor B-ALL. The presence of MRD in day 8 blood and day 29 marrow MRD was associated with shorter event-free survival (EFS) in all risk groups; even patients with 0.01-0.1% day 29 MRD had poor outcome compared to MRD negative patients (59±5% vs 88±1% 5 y EFS). Presence of good prognostic markers TEL-AML1 or trisomies of chromosomes 4 and 10 still provided additional prognostic information, but not in NCI HR patients who were MRD positive. The few patients with detectable MRD at end consolidation fared especially poorly with only a 43±7% 5 y EFS. D29 marrow MRD was the most important prognostic variable in multivariate analysis. The 12% of patients with all favorable risk factors including NCI risk group, genetics, and absence of days 8 and 29 MRD had a 97±1% 5 y EFS with non-intensive therapy. These studies are registered at www.ClinicalTrials.gov as NCT00005585, NCT00005596, and NCT00005603.
INTRODUCTION

The presence of minimal residual disease (MRD) following therapy for acute lymphoblastic leukemia (ALL) has been shown to be an important prognostic marker in many studies \(^1-20\). MRD is typically detected either by polymerase chain reaction (PCR) amplification of clonotypic immunoglobulin or T cell receptor gene rearrangements \(^{20-26}\) or by flow cytometry \(^{27-41}\), the latter based on the principle that leukemic cells express combinations of antigens that are different from those present on normal bone marrow cells. The former technique can be more sensitive, though to achieve adequate sensitivity it is necessary to synthesize optimized clone specific reagents. As a consequence it is difficult to obtain real-time data that could be used for early intervention.

Molecular detection of MRD has been well standardized \(^{25;42-44}\). Though less widely standardized \(^{37;45}\), flow cytometry is faster, generally less expensive, and provides informative results in a higher percentage of patients than molecular methods. For these reasons, flow-based MRD assessment has the potential for rapidly identifying patients at increased risk of relapse, allowing for prompt changes in therapy including earlier intensification \(^7\). Both PCR and flow have successfully been used to help risk-stratify patients and while there is generally concordance between the methods in direct comparisons \(^{46;47}\), individual patients may not always be classified in the same way by each method \(^{48}\).

Although the prognostic significance of MRD in ALL is well established, and is used as a criterion for risk stratification in many current studies \(^{49;50}\), most published studies have
been relatively small. In childhood ALL in particular the value of MRD must be weighed against other well-established prognostic variables, including age, white count, cytogenetic features of blasts, and conventional assessment of response to therapy.\(^{50-57}\). While MRD has been shown to retain prognostic significance after adjusting for some common risk factors,\(^{4;6;19}\) the relationship between MRD and other prognostic factors has been incompletely explored. It is not clear if MRD by itself is all that is needed to predict outcome, if other risk factors add additional information to that obtained by MRD, or whether there are complex interactions between MRD and other factors. For example, we previously showed a difference between the frequency of positive MRD results at end-induction in patients with the two most common favorable genetic lesions: the TEL-AML1 translocation and simultaneous trisomies of chromosomes 4 and 10, which raised the question of whether MRD at end induction has the same significance in both groups.\(^ {28}\)

In 1999 the legacy Pediatric Oncology Group of the Children’s Oncology Group began a prospective study of MRD in all patients enrolled on the classification/induction study P9900. Patients enrolled on this study had MRD measured by flow cytometry at a single central reference laboratory in the peripheral blood (PB) at day 8, and in the bone marrow (BM) at the end of induction (day 29). Blasts from patients with precursor B ALL were analyzed by reverse transcriptase PCR (RT-PCR) and fluorescence in situ hybridization (FISH) methods at a centralized reference laboratory to determine common cytogenetic abnormalities associated with prognosis. Based on results of these studies, and on other clinical and laboratory features, patients were assigned to one of four post-induction treatment protocols. For patients entering the low, standard and high risk protocols,
MRD was again measured in the BM at the end of consolidation. This report describes the relationship of MRD to outcomes for these patients. (Patients entering the very high risk protocol are not reported on here.) Our results demonstrate that end-induction MRD is the single most powerful prognostic marker and that it retains validity in all clinical-and laboratory-defined risk groups. We also show that risk grouping is improved by taking into account MRD assessment performed earlier in therapy as well as other clinical features and genetic characteristics of the leukemia.
MATERIALS AND METHODS

PATIENTS

This study was approved by the IRB of Johns Hopkins University and all participating institutions. All patients were enrolled on P9900 the classification study, which required shipment of PB and BM samples to reference laboratories at the University of New Mexico and the Johns Hopkins Medical Institutions. Patients with precursor-B-cell ALL, as confirmed by the Johns Hopkins laboratory, were eligible for post-induction treatment on therapeutic protocols based on risk group assignment at the end of induction. From 1/2000 through 3/2005, 3686 children were enrolled on the study, among who were 3303 patients with precursor B ALL. Post-induction risk-stratification was based on NCI risk criteria, CNS and testicular status, and on genetic features of their leukemia cells as previously described. Briefly, patients on P9904 included NCI standard risk (SR) patients with a TEL-AML1 translocation or simultaneous trisomies of chromosomes 4 and 10; while patients on P9906 included NCI high risk (HR) patients without favorable genetics who met age, WBC and sex-specific criteria for especially high risk disease originally described by Shuster, or any patient with MLL rearrangement, CNS3 status or testicular involvement. Patients on P9905 included a mixture of NCI SR patients without favorable genetic lesions, NCI HR patients with favorable genetic changes, and other NCI HR patients not meeting the Shuster criteria. Patients <1 year of age, those with Ph+ ALL or hypodiploid ALL, and those failing induction chemotherapy were not eligible for any of these studies and their MRD results are not included here. All patients or their parents or guardians gave separate informed consent for the classification and treatment studies, including consent for measurement of MRD. MRD results were not
made available to physicians caring for the patients, or used to adjust therapy. A total of 827 evaluable precursor B-ALL patients were enrolled on P9904; 1049 on P9905; and 267 on P9906.

Samples were sent to the Johns Hopkins Reference Laboratory for MRD testing from PB at day 8 of induction, and from BM at end of induction (day 29) and end consolidation( weeks 22-30 depending upon the specific protocol). Of the 2143 patients enrolled on 9900 series treatment protocols, samples were submitted on 1946 patients at day 8, 2086 patients at day 29 and 1470 patients at end consolidation. A day 8 BM was also performed and interpreted at local institutions to assess morphologic response to therapy.

MOLECULAR AND CYTOGENETIC TESTING

Ficoll-Hypaque purified BM samples were studied by RT-PCR for the presence of the common translocations E2A-PBX1, TEL-AML1, BCR-ABL and MLL-AF4 as previously described. Other MLL gene rearrangements were detected by FISH studies using a break apart probe strategy. Most patients had conventional cytogenetic studies by COG-certified local laboratories with COG central review of karyotypes. In addition, aliquots were stained with propidium iodide and analyzed by flow cytometry to determine DNA index. All specimens with a DNA index >1.0 had centromeric probe testing for trisomies 4 and 10 as previously described.
THERAPY

Drug doses and the schedule for drug administration are provided in Supplemental tables S1 and S2. Induction therapy was determined by initial white blood cell count and age using NCI criteria and the presence or absence of CNS3 or testicular disease. Patients who were NCI standard risk by age and white count (age >1 yr and < 10; WBC < 50,000/μl) without CNS3 or overt testicular disease were assigned to a three-drug, dexamethasone-based induction; all other patients were assigned to a prednisone-based four-drug induction.

An induction death rate of 1.3% among the first 900 patients led to the closure and amendment of the three drug induction on November 27, 2002. Two changes were made prior to re-opening. The intrathecal therapy on Day 1 was changed from methotrexate to cytarabine and the six doses of native asparaginase at 10,000 IU/m² each were replaced with a single dose of PEG asparaginase. Subsequently there were two deaths from infection among 527 patients (0.38%).

Post-induction therapy on protocols 9904 and 9905 included six courses of intravenous methotrexate with leucovorin rescue with randomized assignment to either one gm/m² as a 24 hour infusion or two gms/m² as a four hour infusion. Patients enrolled on 9904 with the TEL-AML1 translocation with or without trisomies of chromosomes 4 and 10 were also eligible for randomized assignment to receive or not receive a delayed intensification (DI). Patients enrolled on 9905 were eligible for both the MTX randomization and the DI randomization, except that patients with favorable genetics meeting the Shuster high risk
criteria and patients with E2A-PBX1 were assigned to DI and only eligible for the MTX randomization. 9906 was a single regimen study using an “augmented BFM” regimen adapted from Nachman et al.59

FLOW CYTOMETRY

Minimal residual disease was detected by 4-color flow cytometry as previously described28;41. In most cases the two antibody combinations CD20-FITC/CD10-PE/CD45-PerCP/CD19-APC and CD9-FITC/CD34-PE/CD45-PerCP/CD19-PE were sufficient to identify leukemic cells. A minimum of 500,000 events were collected. For day 8 PB analyses, cell numbers frequently precluded acquiring such a number, although in all cases a minimum of 100,000 events were analyzed. When cell numbers were limiting, only the most informative single antibody combination was used for analysis. Sensitivity of .01% was achieved in the great majority of cases. In 4% of cases either cell numbers were limiting or the phenotype overlapped substantially with that of normal B cells so that sensitivity at end of induction was limited to 0.1%; such cases were excluded from analyses that used an MRD threshold of .01%. MRD results were classified as indeterminate in 1.4% of cases, and no results were obtained because no sample was provided in an additional 2.7%, so that overall we obtained satisfactory results at end induction at a sensitivity of at least .01% in 92% of patients enrolled on therapeutic studies.

STATISTICAL ANALYSES
Event-survival estimates were obtained using the Kaplan-Meier method, and standard errors of the estimates were calculated by the method of Peto and Peto. Time to event was calculated as the time from study entry to first event (relapse, secondary malignancy, or death) or date of last contact. The log-rank test was used for comparison of survival curves between various groups. Multivariate analysis was conducted by using Cox proportional hazards regression. Categorical data were compared between groups by using the chi-square or Fisher’s exact test. All tests were conducted at a significance level of 5%. All analyses were carried out using SAS (SAS Institute, Cary, NC) and R (R Foundation for Statistical Computing, Vienna, Austria, http://www.R-project.org) software packages.

RESULTS

End induction MRD is a strong prognostic factor

Figure 1 shows the EFS of 1971 patients enrolled on COG studies P9904, P9905 and P9906 for whom we have MRD results at the end of induction. In general, patients with higher levels of MRD do worse, with those having 0.1-1% MRD having a 49±6% 5 y EFS, and those with >1% MRD having only a 30±8% EFS (not shown). However, even those patients with MRD levels between 0.01% and 0.1% have a much worse outcome (5 y EFS 59±5%) than those with no MRD (EFS 88±1%), suggesting that .01% is an appropriate cutoff for identifying patients at increased risk of relapse. As shown in Table 1, the cutoff of .01%, MRD was prognostic on each study separately, and in both NCI SR and HR patients in study P9905, which included both subsets of patients.
End induction MRD appeared to predict both early (within 3 years) and late relapse. The relapse free survival of patients with and without MRD, with all patients censored at 3 years was used to assess the effect on early relapse, while a similar analysis limited to patients in remission at 3 years was used to assess late relapse (Figure 2). In this analysis (censoring all non-relapse events) we found an early relapse rate of 6.8% among MRD negative patients and 28% among MRD positive (>0.01%) patients (p<.0001). Similarly, the late relapse rate was 4.6% among MRD compared to 24% in MRD positive patients (p<.0001). However, because many patients have not been followed for 3 years, not all late relapses have occurred so that the magnitude of the effect of MRD on late relapse is uncertain.

Interaction of favorable genetic features and end induction MRD

Figure 3 shows separately the outcome of NCI standard risk patients with favorable cytogenetic features (TEL-AML1 translocation or double trisomy(DT) of chromosomes 4 and 10). MRD is strongly prognostic in both subgroups. Although these patients fare very well overall with relatively non-intensive therapy, the proportion of MRD positive patients is much lower in the TEL-AML1+ group than in the DT group (9% vs 19%; p=.0001). The frequency of MRD positivity in the DT patients was similar to that seen in all standard risk patients, and although these patients fared better than other standard risk patients, presence of MRD was still associated with a worse outcome.
Although MRD was an important prognostic marker in patients with and without favorable cytogenetic features, the prognostic significance of cytogenetics depended upon MRD status. Favorable cytogenetic features were associated with better prognosis among patients who were MRD negative, with a 5y EFS of 92±1% for those with favorable genetic features compared to 83±2% for those without (p<.0001). Although TEL-AML1/DT patients who were MRD positive also had a statistically significant better 5 y EFS than those without favorable genetics (57±6% vs. 45±4%, p<.02), their outcome was not that good. The difference in the effect of favorable genetics among MRD positive and negative patients was more readily appreciated when NCI SR and HR patients were looked at separately, and was particularly striking among NCI HR patients. HR MRD negative TEL-AML1 or DT patients had an 83±5% 5 y EFS compared to a 72±4% EFS for MRD negative patients without favorable genetics (p=.004). For HR patients who were MRD positive, however, the 5y EFS was 34±12% for patients with favorable genetics and 33±6% for those without (p=.96). For NCI SR patients, the corresponding figures are 94±1% vs 89±2% for MRD negative patients (p=.0024), and 64±7% vs, 58±6% for MRD positive patients(p=.18).

Prognostic significance of day 8 peripheral blood MRD

The 88% 5 y EFS for patients who were negative in the BM at end of induction, while better than that for those who were positive, does not define a group that needs no further treatment intensification. Moreover, 51% (178/348) of all treatment failures occurred in this group of patients. To try to identify patients with better outcome and/or fewer events,
we examined MRD at an earlier timepoint, specifically in PB at day 8. Overall, 1323/1920 (68.9%) of patients with satisfactory day 8 PB samples were MRD positive at a level of greater than .01%. Figure 4 shows that the presence of day 8 PB MRD was associated with adverse prognosis, and that increasing levels were associated with a progressively poorer outcome. The 5 y EFS of day 8 PB MRD negative patients was 90±2% and only 16% of treatment failures were seen in the 31% of patients who were day 8 PB MRD negative. The 5 y EFS for MRD positive patients was 86±3% for those between .01-.1% MRD; 79±3% for those 0.1-1%; 67±4% for those 1-10% and 54±7% for those>10%.

As shown in Figure 5, patients who had high levels of day 8 PB MRD had a relatively poor outcome even if they cleared their BM of MRD by day 29. Day 29 MRD negative patients with >1% day 8 MRD had a 5 y EFS of 79±4% compared to 90±1% for those with day 8 PB ≤1%. This effect of slow early clearance was most striking among NCI HR patients, where those with >1% day 8 MRD had only a 63±8% 5 y EFS even if they became MRD negative by day 29. (vs 81±3%, p=.0061)

Multivariate analysis of prognostic factors affecting outcome.

Multivariate Cox regression analysis (with main effects only) was employed to determine the relative importance of different prognostic factors that were available at diagnosis, or within the first month of therapy. As shown in Table 2, end induction MRD was the most powerful prognostic factor, followed by NCI risk group, presence of favorable trisomies,
and Day 8 PB MRD. After adjusting for these variables, TEL-AML1 translocation did not achieve statistical significance (p > 0.1). Of note, morphologic assessment of early response, as measured by achieving an M1 BM on day 8, was also not significant.

**Identification of a subgroup of patients with excellent outcome on minimal therapy**

Because several prognostic factors were significant in a multivariate model, we looked for ways to combine these factors to identify patients with an exceptionally good outcome. We found that we could identify patients with a 97±1% 5y EFS achieved with limited therapy. These patients, encompassing approximately 12% of all patients enrolled in these trials were a subset of those enrolled on the P9904 low risk trial—defined by meeting NCI standard risk criteria, without CNS3 or testicular disease, having either DT or TEL-AML1—and an absence of MRD in both the day 8 PB and day 29 BM samples. As shown in Figure 6, no other combination of features identified as favorable a prognostic group. For example, NCI SR patients negative for MRD at both day 8 and day 29, but lacking favorable genetic features, had a 92±3% 5 y EFS (p=.020 vs the best group); while the NCI SR patients with favorable genetics, and absence of day 29 MRD, but with positive day 8 PB MRD had a 93±2% 5y EFS (p=.024).

**Prognostic significance of end consolidation MRD.**

MRD was measured after consolidation therapy, varying from 22-30 weeks after the start of induction therapy depending upon the treatment protocol. (In the case of 9906 the
measurement was at the end of interim maintenance therapy at week 22 but for convenience these will all be referred to as end consolidation MRD.) Although relatively few patients (4.8%) were MRD positive >.01% at this timepoint, presence of MRD was associated with a very poor outcome, with 5 y EFS of 43±7%. (Figure 7) End consolidation MRD was prognostic across studies, with those who were positive having a 5 y EFS of 56±14% on 9904(n=14); 46±11% on 9905 (n=29); and 25±15% on 9906 (n=15). End consolidation MRD was also prognostic in multivariate analysis, with a hazard ratio of 2.25 (p=.0002), entering the model in a stepwise regression after day 29 MRD and NCI risk group (not shown).
DISCUSSION

The presence of MRD has been shown to be a very strong predictor of relapse in both children and adults with ALL.\(^1\)\(^{-20}\) This has been true across studies employing different analytical methods and treatment strategies. In almost all studies, MRD has been shown to be prognostic at essentially every time point studied, though the most useful measurements appear to be relatively early in therapy, during or after induction and early in consolidation\(^2\)\(^;4\)\(^{-6}\)\(^;9\)\(^;15\)\(^;16\)\(^;19\). However many additional variables are associated with outcome, especially in childhood ALL\(^50\)\(^;57\). While some studies have been able to demonstrate that MRD is prognostically important even after adjusting for other risk factors\(^4\)\(^;6\)\(^;19\), or in patients of a defined risk group,\(^1\)\(^;20\)\(^;63\)\(^;64\) most studies have been too small to be able to investigate fully the relationship between MRD and other prognostic factors in outcome prediction.

We evaluated the prognostic significance of MRD determined via flow cytometry in nearly 2000 children with ALL entered on the COG classification study P9900, in which blast cell cytogenetics were determined uniformly for purposes of risk assignment. This sample size allowed us to determine the prognostic significance of MRD in the context of therapies of varied intensity, and also to explore the interaction of MRD with other known risk factors. While these protocols had randomizations to different treatments, data are presented here without respect to the specific therapeutic regimens. However, there was no difference in the proportion of MRD positive and negative patients assigned
to each of the treatment arms so it is unlikely that any of our conclusions regarding MRD and outcome are due to differences in post-induction therapy.

Importantly, the methodology we employed is robust, as informative day 29 MRD data with a sensitivity of at least 0.01% were generated for 92% of patients overall, and 94% of those for whom specimens were submitted to the central reference laboratory. Results were available within 24 hours of sample receipt. In the current generation of COG ALL trials we use the end-induction MRD value to alter the intensity of post-induction therapy and data are available for real-time clinical interventions in over 95% of over 5000 patients enrolled to date.

MRD was a strong predictor of outcome in all analyses. In general, higher levels of MRD were associated with increasingly poor outcome, but even those patients with as little as 0.01–0.1% had worse event free survival, suggesting that .01% is the most appropriate cutoff to identify higher-risk patients for potential intervention. While patients at increasingly poor risk by conventional characteristics had a progressively worse outcome irrespective of whether they were MRD positive or not (see Table 1), those who had poor clinical and laboratory risk factors in addition to MRD had the worst outcome. In particular NCI HR patients who were day 29 MRD positive had a 5 y EFS of less than 35% irrespective of whether they were treated on P9905 or P9906.

To our knowledge, this is the first study to demonstrate the provocative result that end induction MRD appears to predict late as well as early relapse. Although the mechanism
of early and late relapse appears to be different\textsuperscript{65}, in both cases we detect resistant tumor cells early in therapy. There were proportionately approximately 4 times as many early relapses and 5 times as many late relapses in the MRD positive group than in the MRD negative group. While the magnitude of the latter effect may not be accurately estimated (because not all late relapses have yet occurred) it is essentially impossible that enough additional relapses could occur to affect this conclusion.

Day 29 MRD was the most important prognostic factor in a Cox stepwise linear regression analysis. MRD was prognostic across all risk groups, including those with favorable genetic features. The higher rate of MRD in patients with favorable trisomies compared to those with TEL-AML1 translocations had led us previously to suggest that the prognostic significance of early MRD might not be seen in the former because of intrinsically slower clearance of the disease\textsuperscript{28}. However, this is not true; MRD was still an important risk factor in these patients. MRD did not, however, fully account for outcome. When adjusted for MRD, blast genetic features were still a significant prognostic factor, though the relationships were complex. In patients who were MRD negative, favorable genetic features were associated with better prognosis, both among NCI standard and NCI high risk patients. However, particularly in HR patients, favorable cytogenetic features were not sufficient to overcome the adverse prognosis of MRD.

In our series, about half of events occurred among patients who were MRD negative at the end of induction. This contrasts with the results of many studies that employ molecular methods, which find very few failures among patients who are MRD negative
at end induction. This may in part reflect the greater sensitivity of molecular MRD testing. Other studies that have used flow methods have seen a relatively higher proportion of events in patients who were MRD negative at end of induction. It is also possible that we missed some MRD positive patients because our method only detects viable cells and small numbers of residual tumor cells may not survive shipping. However, it should also be noted that the intensity of therapy given to many of our patients was lower than in most other series, and the prognostic importance of MRD has been shown to depend on therapeutic regimen. It is thus possible that with different therapy our MRD negative cohort would have a better outcome.

While detection of MRD at end of induction helps to identify patients at high risk of relapse, it did not do as well at identifying patients at low risk. To address this, we turned to an earlier time, measuring MRD in the PB at day 8. While PB MRD has been shown not to be equivalent to BM MRD in precursor B ALL, the prognostic significance of PB MRD itself has not been established. We found that, as with day 29 BM, the presence of MRD in PB at day 8 is associated with adverse outcome, with an increasingly bad outcome seen with progressively higher levels. Most importantly, only 16% of events were seen among patients who were day 8 MRD negative, indicating that this is an even better prognostic group than those who were MRD negative in the BM at day 29.

However, in order to identify the best group of patients, it was necessary to combine MRD response with other patient characteristics. The patients with the best outcome were NCI SR patients who were MRD negative at both day 8 and day 29, and whose blasts had favorable cytogenetic features. Only 7 patients of 261 have had treatment
failures (only 5 of which were relapses); this accounts for only 2% of the total failures in the series. This outstanding outcome was attained with relatively limited antimetabolite-based chemotherapy. All DT patients and half of the TEL-AML1+ patients received this therapy that included no delayed intensification phase, no anthracyclines, and no alkylating agents. The remaining one quarter of patients (half the TEL-AML1+ group) received a single delayed intensification phase. Thus, MRD can identify a cohort of 10-15% of B-precursor ALL patients for whom the least toxic therapy may be most appropriate. These patients will not benefit from more intensive therapy, and can avoid the risks associated with treatment intensification.

The value of looking at clearance of leukemic blasts earlier in therapy has been recognized by others as well. Coustan-Smith et al demonstrated a 6% cumulative relapse rate among 51 patients with <.01% blasts by flow cytometry at day 19 of therapy. In another study using molecular methods of detection, there were no failures seen in 14 patients who had <10^{-4} blasts by day 15. Day 15 MRD has also been suggested as a means to identify patients at especially high risk and indeed our data also suggest that by using flow cytometry with a high cutoff for positivity, it is possible to identify a poor risk group of patients as early as day 8. The 5 year EFS of patients with >10% leukemic cells in the PB at day 8 was only 54%, and those with >1% MRD who cleared their BMs by day 29 had only a 79% 5 y EFS (and only 63% among NCI HR patients).

The patients with the poorest outcome were those who had persistent MRD at the end of consolidation therapy. This time point varied from 22-30 weeks depending upon the
specific protocol. Not surprisingly, the highest risk patients treated on 9906 had the worst outcome, with a 25% 5 y EFS. However, it is more surprising that end consolidation positive patients did as well as they did; in spite of the fact that these patients only got maintenance chemotherapy from then on, the overall 5y EFS was >40%. Most studies that looked at a second post-induction time-point have looked earlier in therapy than we have, so direct comparison with those results is difficult. In general patients with persistent MRD long into therapy have a very poor outcome, but in most series, relatively small numbers of patients continue to be positive this late. In addition the ability to rescue such patients has been called into question so that the practical value of identifying such patients is limited.

It is interesting to contemplate how our results might be used to refine risk grouping for treatment purposes. One of the advantages of flow cytometry over sensitive molecular techniques for detecting MRD is that results are potentially available very early in therapy, and changes can be made shortly after, or conceivably even during induction. Early intensification of therapy has been shown to help patients with slow response to therapy assessed by BM morphology. In current COG studies patients with >0.1% end induction MRD are assigned to augmented therapy during the immediate post-induction period; this threshold was chosen before the outcome results described here were available and was deliberately designed to be conservative, based on published data that most closely reflected the methods and therapy currently used by COG. While these patients do in fact have poor outcome, results of the current study suggest that more intensive intervention may be desirable for all patients with MRD>0.01% at end of
induction. Whether this intervention should be the same irrespective of other risk factors is not yet clear, especially because the results presented here were obtained with many patients receiving less intensive therapy than is used at the current time. As noted above, absence of day 8 PB MRD, combined with other favorable factors, can be used to identify a group of patients in whom to consider more limited, or even de-escalation of therapy. Higher levels of day 8 MRD could also be used to identify additional patients for intensification, possibly even during induction. Most significantly, however, by combining day 8 PB and day 29 BM MRD, there is no longer any prognostic significance to morphologic assessment of BM at day 8, indicating this conventional measure of early response\(^55\) may not be necessary in the future.

In summary, we have demonstrated that MRD after induction therapy is the most important prognostic factor for outcome in children with ALL. Measurement of MRD in the PB at day 8 provides additional useful information, especially to help identify patients at low risk of relapse, and that with these two measurements of response, day 8 BMs are no longer necessary to assess response. However, additional clinical and laboratory features, especially the presence of favorable cytogenetic abnormalities among patients who are MRD negative, do provide additional prognostic information; by combining all favorable good risk features it is possible to identify a group of patients that accounts for about 12% of all children with B-precursor ALL who are almost certain to be cured with limited therapy that does not include any anthracyclines or alkylating agents.
Acknowledgments

The authors are grateful to Karen Bowles and Shirley Fuller for excellent technical assistance with the MRD studies, and to the physicians and CRAs at the COG institutions participating in this study for cooperation with the sample submission requirements. The authors declare no competing financial interests.

Supported by Grants R01 CA86011 and U10 CA98543 from NCI

Author contributions

Borowitz: Designed research, collected data, analyzed and interpreted data, drafted the manuscript
Devidas: analyzed and interpreted data, performed statistical analysis
Hunger: analyzed and interpreted data, drafted the manuscript
Bowman: collected data, analyzed and interpreted data
A. Carroll: collected data
W. Carroll: analyzed and interpreted data
Linda: performed statistical analysis
Martin: collected data, analyzed and interpreted data
Pullen: designed research, collected data, analyzed and interpreted data
Viswanatha: collected data
Willman: collected data
Winick: collected data, analyzed and interpreted data
Camitta: Designed research; analyzed and interpreted data; drafted the manuscript
Figure Legends

Figure 1. Event-free survival of all patients enrolled on 9900 series therapeutic studies with satisfactory end of induction MRD. 5 y EFS values are shown for patients with varying levels of MRD. The outcome of those with high levels of MRD is very poor but even those with 0.01-0.1% MRD have only a 59±5% 5 y EFS.

Figure 2. Relapse-free survival showing the effect of end induction MRD on early (A) and late (B) relapse. MRD positivity is defined as >.01%. In A, all patients were censored at 3 years from diagnosis. In addition all non-relapse events occurring during the first three years were censored. In B, only patients who were in remission at 3 years from diagnosis are included in the analysis and again all non-relapse events are censored.

Figure 3. Event-free survival of NCI SR patients with favorable genetic features (A: TEL-AML1; B:Double trisomies. The very few patients with both lesions are included in A) as a function of end induction MRD. Outcome of MRD positive patients in both groups is much worse than those who are negative. 5 y EFS is indicated on each curve as appropriate.

Figure 4. Event free survival of all patients as a function of level of day 8 peripheral blood MRD. There is a stepwise decrement in 5 y EFS at each 10 fold increase in MRD level.
Figure 5. Event free survival among day 29 MRD negative patients as a function of day 8 blood MRD. Patients with high levels (defined as >1%) MRD at day 8 fare worse (5 y EFS of 79±4%) than those with lower levels (90±1%) even if they become MRD negative by day 29. This difference was especially important in NCI HR patients (see text).

Figure 6. Event-free survival among variably defined groups of good prognosis patients. NCI SR patients with favorable genetic features who were MRD negative at both day 8 and day 29 were the best group, with a 97±1% 5 y EFS. They had statistically better outcomes than either patients without the genetic features who had the same MRD characteristics (92±3%; p=.020) or end induction MRD negative, favorable genetic patients who were day 8 MRD positive (93±2%; p=.024).

Figure 7. Event free survival of all patients as a function of level of end of consolidation MRD. Patients who were MRD positive (>0.1%) had a significantly inferior outcome, with a 5 y EFS of 43±7%. This effect was seen on each of the three therapeutic studies (see text).
Fig 1. Prognostic significance of end of induction MRD in all patients enrolled on therapeutic studies.

EFS by Day 29 MRD

- MRD negative (≤ 0.01%) (n = 1588)
- 0.01% < MRD ≤ 0.1% (n = 175)
- 0.1% < MRD ≤ 1.0% (n = 141)
- MRD > 1.0% (n = 67)

P < 0.0001
Fig 2. Prognostic significance of end induction MRD on early vs. late relapse.

9904+9905+9906
Relapse-free survival by Day 29 MRD

Early relapse
Time to failure censored at 3 y

Late relapse
Time to failure for pts in remission at 3 y

A

B

Relapse-free survival probability

Years

Negative (n = 1588)
Positive (n = 363)

P < 0.0001

Relapse-free survival probability

Years

Negative (n = 1240)
Positive (n = 233)

P < 0.0001
Fig 3. Prognostic significance of end of induction MRD in NCI standard risk patients with favorable genetic features.
Fig 4. Prognostic significance of day 8 blood MRD in all patients enrolled on therapeutic studies.

**EFS by Day 8 MRD**

- **MRD Negative (≤ 0.01%)** (n = 597)
- **0.01% < MRD ≤ 0.1%** (n = 339)
- **0.1% < MRD ≤ 1.0%** (n = 497)
- **1.0% < MRD ≤ 10.0%** (n = 372)
- **MRD > 10%** (n = 115)

*P < 0.0001*
Fig 5. Prognostic significance of day 8 blood MRD in patients who are free of MRD in bone marrow by day 29.

<table>
<thead>
<tr>
<th>Event-free survival probability</th>
<th>Years</th>
</tr>
</thead>
<tbody>
<tr>
<td>90±1%</td>
<td>0</td>
</tr>
<tr>
<td>79±4%</td>
<td>7</td>
</tr>
</tbody>
</table>

$P < 0.0001$

- Day 8 MRD negative ($\leq 1.0\%$) ($n = 1174$)
- Day 8 MRD positive ($> 1.0\%$) ($n = 269$)

$9904 + 9905 + 9906$

Day 29 Marrow MRD Negative (Cutoff=0.01%)
Fig 6. Identification of a subgroup of patients with excellent outcome.

**9904/9905**

**EFS by miscellaneous Day-8/Day-29 MRD categories**

- **9904**: Day 8 \( \leq 0.01\% \), Day 29 \( \leq 0.01\% \) (\( n = 261 \))
- **9905 SR no TEL/Tris**: Day 8 \( \leq 0.01\% \), Day 29 \( \leq 0.01\% \) (\( n = 143 \))
- **9904**: Day 8 > 0.01\%, Day 29 \( \leq 0.01\% \) (\( n = 339 \))

\[ P = 0.0465 \]
Fig 7. Prognostic significance of end of consolidation MRD in all patients enrolled on therapeutic studies.

9904 + 9905 + 9906
EFS by end of consolidation MRD

Event-free survival probability

Years

9904 + 9905 + 9906

83±1%

43±7%

$P < 0.0001$

- Negative ($\leq 0.01\%$) ($n = 1161$)
- Positive ($> 0.01\%$) ($n = 58$)
Table 1

Effect of End Induction MRD on Outcome by Therapeutic Protocol and NCI Risk Group

<table>
<thead>
<tr>
<th>STUDY/RISK GROUP</th>
<th>MRD negative</th>
<th>MRD positive (&gt; .01%)*</th>
</tr>
</thead>
<tbody>
<tr>
<td>9904 (SR)</td>
<td>95±1 (668)</td>
<td>64±7 (105)</td>
</tr>
<tr>
<td>9905 SR</td>
<td>89±2 (464)</td>
<td>59±6 (119)</td>
</tr>
<tr>
<td>9905 HR</td>
<td>79±4 (295)</td>
<td>33±8 (74)</td>
</tr>
<tr>
<td>9906 (HR**)</td>
<td>72±5 (161)</td>
<td>34±8 (85)</td>
</tr>
</tbody>
</table>

*p<.0001 for all comparisons
** Also includes CNS3 patients, patients with testicular disease, or patients with MLL rearrangements

Table 2

Cox Multivariate Analysis.
All variables shown entered in model.

<table>
<thead>
<tr>
<th>VARIABLE</th>
<th>HAZARD RATIO</th>
<th>P VALUE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Day 29 MRD&gt;.01%</td>
<td>4.31</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>NCI Risk Group</td>
<td>2.25</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Trisomy 4 and 10</td>
<td>.570</td>
<td>.0005</td>
</tr>
<tr>
<td>Day 8 MRD (PB)&gt; .01%</td>
<td>1.51</td>
<td>.0179</td>
</tr>
<tr>
<td>TEL-AML1</td>
<td>.778</td>
<td>.1506</td>
</tr>
<tr>
<td>Day 8 M1 marrow</td>
<td>1.034</td>
<td>.7890</td>
</tr>
</tbody>
</table>
Reference List


42. Gabert J, Beillard E, Van Der Velden V et al. Standardization and quality control studies of 'real-time' quantitative reverse transcriptase polymerase chain reaction of fusion gene transcripts for residual disease detection in leukemia - a Europe Against Cancer program. Leukemia 2003;17:2318-2357.


52. Shurtleff SA, Bujis A, Behm FG et al. TEL/AML1 fusion resulting from a cryptic (12;21) is the most common genetic lesion in pediatric ALL and defines a subgroup of patients with an excellent prognosis. Leukemia 1995;9:1985-1989.


68. Van Der Velden V, Jacobs DC, Wijkhuijs AJ et al. Minimal residual disease levels in bone marrow and peripheral blood are comparable in children with T cell acute lymphoblastic leukemia (ALL), but not in precursor-B-ALL. Leukemia 2002;16:1432-1436.
