Regular Article

CLINICAL TRIALS AND OBSERVATIONS

Prognostic significance of minimal residual disease in high risk B-ALL: a report from Children's Oncology Group study AALL0232

Michael J. Borowitz,[1] Brent L. Wood,[2] Meenakshi Devidas,[3] Mignon L. Loh,[4] Elizabeth A. Raetz,[5] Wanda L. Salzer,[6] James B. Nachman,[7] Andrew J. Carroll,[8] Nyla A. Heerema,[9] Julie M. Gastier-Foster,[10] Cheryl L. Willman,[11] Yunfeng Dai,[3] Naomi J. Winick,[12] Stephen P. Hunger,[13] William L. Carroll,[14] and Eric Larsen[15]

[1]Department of Pathology, Johns Hopkins Medical Institutions, Baltimore, MD; [2]Department of Laboratory Medicine, University of Washington, Seattle, WA; [3]Department of Biostatistics, Colleges of Medicine and Public Health & Health Professions, University of Florida, Gainesville, FL; [4]Benioff Children's Hospital and the Helen Diller Family Comprehensive Cancer Center, University of California, San Francisco, San Francisco, CA; [5]Department of Pediatrics, University of Utah, Salt Lake City, UT; [6]US Army Medical Research and Materiel Command, Fort Detrick, MD; [7]Department of Pediatrics, University of Chicago, Chicago, IL; [8]Department of Genetics, University of Alabama at Birmingham, Birmingham, AL; [9]Department of Pathology, The Ohio State University School of Medicine, Columbus, OH; [10]Nationwide Children's Hospital, Columbus, OH; [11]Cancer Center and Departments of Internal Medicine and Pathology, University of New Mexico, Albuquerque, NM; [12]University of Texas Southwestern Medical Center, Dallas, TX; [13]Department of Pediatrics and the Center for Childhood Cancer Research, Children's Hospital of Philadelphia, Philadelphia, PA; [14]Department of Pediatrics and The New York University Cancer Institute, New York University Medical Center, New York, NY; and [15]Maine Children's Cancer Program, Scarborough, ME

Key Points

- MRD measured by flow cytometry is prognostic in childhood B-ALL even with more effective high-dose methotrexate therapy.

- Intensive therapy in MRD-positive patients altered the timing of relapse but did not overcome the poor prognostic significance of MRD.

Minimal residual disease (MRD) is highly prognostic in pediatric B-precursor acute lymphoblastic leukemia (B-ALL). In Children's Oncology Group high-risk B-ALL study AALL0232, we investigated MRD in subjects randomized in a 2 × 2 factorial design to receive either high-dose methotrexate (HD-MTX) or Capizzi methotrexate (C-MTX) during interim maintenance (IM) or prednisone or dexamethasone during induction. Subjects with end-induction MRD ≥0.1% or those with morphologic slow early response were nonrandomly assigned to receive a second IM and delayed intensification phase. MRD was measured by 6-color flow cytometry in 1 of 2 reference labs, with excellent agreement between the two. Subjects with end-induction MRD <0.01% had a 5-year event-free survival (EFS) of 87% ± 1% vs 74% ± 4% for those with MRD 0.01% to 0.1%; increasing MRD amounts was associated with progressively worse outcome. Subjects converting from MRD positive to negative by end consolidation had a relatively favorable 79% ± 5% 5-year disease-free survival vs 39% ± 7% for those with MRD ≥0.01%. Although HD-MTX was superior to C-MTX, MRD retained prognostic significance in both groups (86% ± 2% vs 58% ± 4% for MRD-negative vs positive C-MTX subjects; 88% ± 2% vs 68% ± 4% for HD-MTX subjects). Intensified therapy given to subjects with MRD >0.1% did not improve either 5-year EFS or overall survival (OS). However, these subjects showed an early relapse rate similar to that seen in MRD-negative ones, with EFS/OS curves for patients with 0.1% to 1% MRD crossing those with 0.01% to 0.1% MRD at 3 and 4 years, thus suggesting that the intensified therapy altered the disease course of MRD-positive subjects. Additional interventions targeted at the MRD-positive group may further improve outcome. This trial was registered at www.clinicaltrials.gov as #NCT00075725. (*Blood*. 2015;126(8):964-971)

Introduction

Minimal residual disease (MRD) is highly predictive of relapse in children, adolescents, and young adults treated for acute lymphoblastic leukemia (ALL).[1-6] MRD is typically measured either by assessment of clone-specific markers of immunoglobulin and/or T-cell receptor gene rearrangements using polymerase chain reaction (PCR) or by flow cytometry, taking advantage of the fact that leukemic cells have phenotypes that allow them to be distinguished from normal cells. The Children's Oncology Group (COG) has been assessing MRD by flow cytometry at the end of 4 weeks of induction therapy in subjects with newly diagnosed ALL since 1999, and has previously demonstrated

that this is the most powerful predictor of outcome in children, adolescents, and young adults with B-precursor ALL (B-ALL).[1]

Because MRD is known to be such a strong prognostic factor, most studies of childhood ALL use this to determine risk stratification, which determines the intensity of postinduction therapy. Nearly all studies have nonrandomly assigned MRD-positive subjects to more intensive therapeutic arms. The Associazione Italiana di Ematologia Oncologia Pediatrica and the Berlin-Frankfurt-Munster ALL 2000 study used PCR-based MRD assays performed at day 33 and day 78 to divide subjects into low-, intermediate-, and high-risk groups.[7] St Jude

Children's Research Hospital employs flow cytometric assessment of MRD to identify high-risk subjects early in therapy based on MRD positivity at day 42.[8,9] The Medical Research Council United Kingdom ALL2003 trial group randomized subjects with end-induction MRD (using PCR-based methods) ≥0.01% to receive more or less intensive therapy.[10] Many other pediatric ALL trials incorporate MRD into their treatment strategies.[11-15]

Though our initial studies only assessed the prognostic significance of MRD without intervention, in 2003, we began using MRD as one variable to determine the intensity of postinduction therapy. Here, we report the effect of MRD on the outcome of subjects with NCI high-risk (HR) ALL treated on COG trial AALL0232 (NCT00075725). This protocol was designed to assess the effect of induction with 14 days of dexamethasone (Dex) vs 28 days of prednisone (Pred) and also to compare effectiveness of a 2-month interim maintenance (IM) block using high-dose methotrexate (HD-MTX) plus leucovorin rescue versus escalating IV MTX (without rescue) plus pegaspargase (Capizzi methotrexate [C-MTX]). Subjects with a slow early response (SER), defined as either bone marrow (BM) MRD ≥0.1% at day 29 of induction therapy or BM blasts ≥5% by morphology at day 15 of induction therapy, were included in randomizations but were nonrandomly assigned to receive second C-MTX IM and delayed intensification (DI) phases prior to starting maintenance therapy and prophylactic cranial radiotherapy. Although our prior studies showed prognostic significance of MRD down to a level of 0.01%, these results were not available at the time AALL0232 was designed, and the MRD-based definition of SER in this trial used a threshold of ≥0.1%.

Figure 1. Schematic of study design for treatment on AALL0232. See text for details.

Materials and methods

All samples were derived from patients age 1 to 30 years with NCI HR (age 10+ years and/or initial white blood cell count ≥50 000/μL) precursor B ALL enrolled on COG AALL0232. The diagnosis was determined by routine morphology and immunophenotyping. AALL0232 enrolled 3154 subjects between January 2004 and January 2011 with 2914 eligible subjects randomized prior to the start of induction therapy. MRD was successfully measured at the end of induction in 2479 of the eligible subjects who form the basis of this report; MRD results were available for postinduction risk stratification in all but 18 otherwise-eligible patients. All subjects provided informed consent for these studies, in accordance with the Declaration of Helsinki. The laboratory studies were performed under a protocol approved by the Johns Hopkins University School of Medicine (JHU) or the University of Washington (UW) institutional review board. This clinical trial was registered at www.clinicaltrials.gov as #NCT00075725.

Subjects were treated using a COG augmented Berlin-Frankfurt-Munster–based chemotherapy backbone with two randomizations in a 2X2 factorial design described in more detail elsewhere. Figure 1 shows an abbreviated schema of the study design, and a detailed description is shown in supplemental Figure 1 (available on the *Blood* Web site). The study compared 14 days of Dex to 28 days of Pred in a 4-drug induction regimen that also included vincristine, daunorubicin, intrathecal chemotherapy and pegaspargase. AALL0232 also compared C-MTX to HD-MTX during the first 8-week IM phase. Subjects were classified as rapid or slow early responders as described above. Subjects with a t(9;22) and/or *BCR-ABL1* fusion were not eligible to continue on study postinduction and are excluded from these analyses. All subjects received an identical 8-week augmented Berlin-Frankfurt-Munster consolidation. Prior to maintenance therapy, rapid early responders with ≤5% BM blasts at day 15 and ≤0.1% MRD at day 29 received a single IM and DI phase, while SERs with either BM blasts ≥5% on day 15 or BM MRD ≥0.1% at day 29 received a second IM with C-MTX and a second DI phase along with prophylactic cranial radiation (1200 cGy). Subjects with BM blasts 5% to 25% and/or MRD ≥1% at end of induction were given an additional 2 weeks of induction therapy. If the day 43 BM had ≥5% blasts or MRD was >1%, they were removed from protocol therapy and treated at physician discretion with outcome data captured.
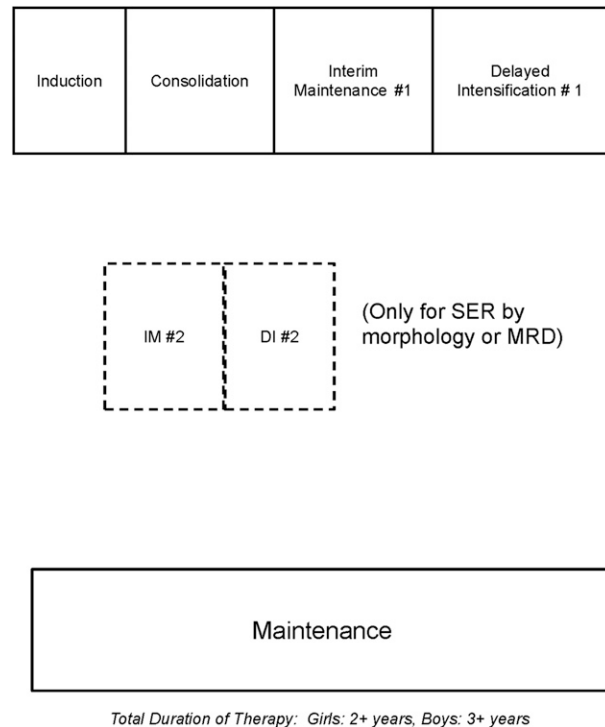
Otherwise, they also received the same postinduction therapy given to SER subjects. AALL0232 also requested but did not require an MRD sample at the end of the consolidation phase for subjects assigned to receive 2 IM and DI phases. These MRD samples were obtained in 197 out of 339 MRD-positive subjects (58%) who had not had an event before end consolidation.

Flow cytometric detection of MRD

End-induction BM samples were submitted to either UW or JHU based on geography. MRD was detected at both sites using 6-color flow cytometry using a modification of our previously described method.[1,16] Samples were stained with 2 different 6-color antibody combinations (CD20-FITC/CD10-PE/CD38-PerCPCy5.5/CD58-APC/CD19-PECy7/CD45-APCH7 and CD9/CD13+33/CD34/CD10/CD19/CD45). A third tube contained SYTO-16 to identify all nucleated cells using a method previously described by Dworzak et al.[17] CD19 in this tube was used to express B cells as a percent of all nucleated cells; MRD identified in either of the two test tubes was expressed as a percent of B cells and the third tube used to calculate MRD as a percent of nucleated cells. Finally, as in our prior studies,[1,18] mononuclear cells were estimated on a display of CD45/SSC to exclude granulocytes and MRD ultimately expressed as a percent of mononuclear cells. All antibodies were obtained from BD Biosciences or BD Pharmingen (San Jose, CA; San Diego, CA) except for CD19 and CD58, which came from Beckman Coulter (Miami, FL). Samples were acquired on either a FACSCanto flow cytometer (JHU) or a modified LSRII (UW) and analyzed with software written by one of us (B.L.W.). A minimum of 750 000 events was acquired in each of the 2 tubes. MRD was identified based on the position of cells on dual parameter displays in areas known not to contain any normal elements (so-called empty space), based on our prior studies of normal and regenerating marrows.[16] Sensitivity of detection was in part a function of the phenotype of the leukemic cells and of the number of background normal B-cell precursors but was at least 0.01% in >95% of cases. In some cases, we were able to detect MRD at a level below this threshold, but except where specifically indicated, such cases were considered negative for the purposes of this analysis. An example of high-sensitivity MRD detection, even in the presence of normal B-cell precursors, is illustrated in supplemental Figure 2.

Table 1. Detailed description of the study population

Variable	Overall (n = 2473)	MRD at day 29				
		<0.01% (n = 1788)	0.01% ≤ MRD < 0.10% (n = 281)	0.10% ≤ MRD < 1.0% (n = 230)	1.0% ≤ MRD < 10.0% (n = 123)	≥10.0% (n = 51)
Age, mean (SD)	10.5 (5.7)	9.8 (5.5)	11.3 (6.1)	12.4 (5.3)	13.1 (5.2)	13.9 (4.6)
Age group (y)						
<10	823 (33.3%)	663 (37.1%)	81 (28.8%)	48 (20.9%)	26 (21.1%)	5 (9.8%)
≥10	1650 (66.7%)	1125 (62.9%)	200 (71.2%)	182 (79.1%)	97 (78.9%)	46 (90.2%)
Gender						
Male	1350 (54.6%)	944 (52.8%)	164 (58.4%)	137 (59.6%)	71 (57.7%)	34 (66.7%)
Female	1123 (45.4%)	844 (47.2%)	117 (41.6%)	93 (40.4%)	52 (42.3%)	17 (33.3%)
Treatment						
HD-MTX	1237 (50.0%)	915 (51.2%)	132 (47.0%)	109 (47.4%)	59 (48.0%)	22 (43.1%)
C-MTX	1236 (50.0%)	873 (48.8%)	149 (53.0%)	121 (52.6%)	64 (52.0%)	29 (56.9%)
WBC ($\times 10^9$/L), median (min, max)	25 (0.3, 1306)	25.3 (0.3, 1306)	27.4 (0.8, 495.5)	20.8 (0.4, 900)	18.4 (1, 616)	62.8 (0.9, 845)
WBC						
<50 × 10^9/L	1398 (56.5%)	1007 (56.3%)	158 (56.2%)	141 (61.3%)	71 (57.7%)	21 (41.2%)
≥50 × 10^9/L	1075 (43.5%)	781 (43.7%)	123 (43.8%)	89 (38.7%)	52 (42.3%)	30 (58.8%)
CNS status						
CNS1	2129 (86.1%)	1547 (86.5%)	237 (84.3%)	199 (86.9%)	103 (83.7%)	43 (84.3%)
CNS2	343 (13.9%)	241 (13.5%)	44 (15.7%)	30 (13.1%)	20 (16.3%)	8 (15.7%)
TEL-AML1	331 (15.6%)	291 (18.9%)	29 (11.9%)	6 (3.1%)	5 (4.8%)	
MLL	58 (2.8%)	41 (2.7%)	15 (6.4%)			2 (5.1%)
Trisomy 4 and 10	360 (17.2%)	269 (17.8%)	49 (20.3%)	29 (14.9%)	10 (9.6%)	3 (7.9%)
Day 15 morphology (SER)	265 (10.7%)	62 (3.5%)	40 (14.2%)	59 (25.7%)	60 (48.8%)	44 (86.3%)

AML, acute myeloid leukemia; CNS, central nervous system; MLL, mixed-lineage leukemia; WBC, white blood cell.

MRD results obtained from the 2 laboratories were highly comparable. As shown in supplemental Figure 3, the frequency of detection of MRD at any level of positivity was virtually identical between laboratories. In addition, a limited sample exchange between laboratories showed highly comparable results (supplemental Figure 4).

Statistical methods

Event-free survival (EFS) was the primary outcome for most analyses; events included relapse, induction failure (≥25% BM blasts at day 29), death from any cause, or development of a second malignancy. In the primary analyses, subjects who were removed from the study because of persistent BM blasts ≥5% or MRD >1% after day 43 were considered as events. Overall survival (OS) was performed using time to death by any cause as an event. Disease-free survival (DFS) was used as the outcome measure to assess the effect of end-consolidation MRD, because only those MRD-positive subjects receiving postinduction therapy on protocol were eligible to have a sample assayed. Figures display estimates of EFS, OS, or DFS rates computed using the Kaplan-Meier method for the various subgroups. The log-rank test was used to compare survival curves. Cox proportional hazards model was used for multivariate analysis of outcomes adjusting for clinical and biological characteristics. Proportions were compared between groups using the χ^2 test. Table 1 presents descriptive statistics for all patients enrolled in the study.

Table 2. Summary of all events comparing MRD-positive and negative patients

Event	All (n = 2473)	MRD <0.01% (n = 1788)	MRD ≥0.01% (n = 685)
Induction failures	25	0	25
Relapses	355	162	193
Deaths	267	119	148
Induction deaths	13	9	4
Protocol-therapy–related deaths	37	28	9
Second malignancies	22	18	4

Results

Table 2 shows a summary of events for children enrolled in the protocol as a function of MRD status. There were a total of 355 relapses, with significantly more occurring in the MRD-positive group. Thirteen subjects died in induction, and there were 37 deaths attributed to protocol therapy and 22 to second malignancies. There was no difference in these latter 3 categories between MRD-positive and MRD-negative patients. Thirty-seven (9.2%) of subjects were withdrawn from the study in first complete remission (CR1) to pursue transplant; there were no relapses in these subjects prior to or during the follow-up period.

Effect of end-induction MRD on outcome

Figure 2A shows the EFS of subjects enrolled on AALL0232 as a function of MRD levels at the end of induction separated into deciles. Increasing levels of MRD were significantly associated with worsening outcome. Curiously, subjects who had MRD levels between 0.1% and 1% had an unusual shape to their survival curve. For about 18 months, the outcome of these subjects approximated those who were MRD <0.01%, but then dropped off, crossing the curve of those with MRD 0.01% to 0.1% at ~3 years. Of 77 total relapses in this cohort, only 18 (23%) had occurred by year 2 and 40 (52%) by year 3. By contrast, 52% of relapses of subjects MRD positive between .01% and 0.1% occurred by year 2 and 67% by year 3; for MRD <0.01% subjects, 48% of relapses occurred in the first 2 years and 72% by year 3. Recall that these subjects with 0.1% to 1% MRD and the lower early relapse rate received a second IM and DI phase not given to these with MRD <0.1%, which clearly influenced the shape of the EFS curves. Subjects with >1% MRD were also offered more intensive therapy, and those with 1% to 10% MRD showed a smaller, less dramatic inflection point, indicating a lesser (if any) effect of intensification in this higher-risk group.
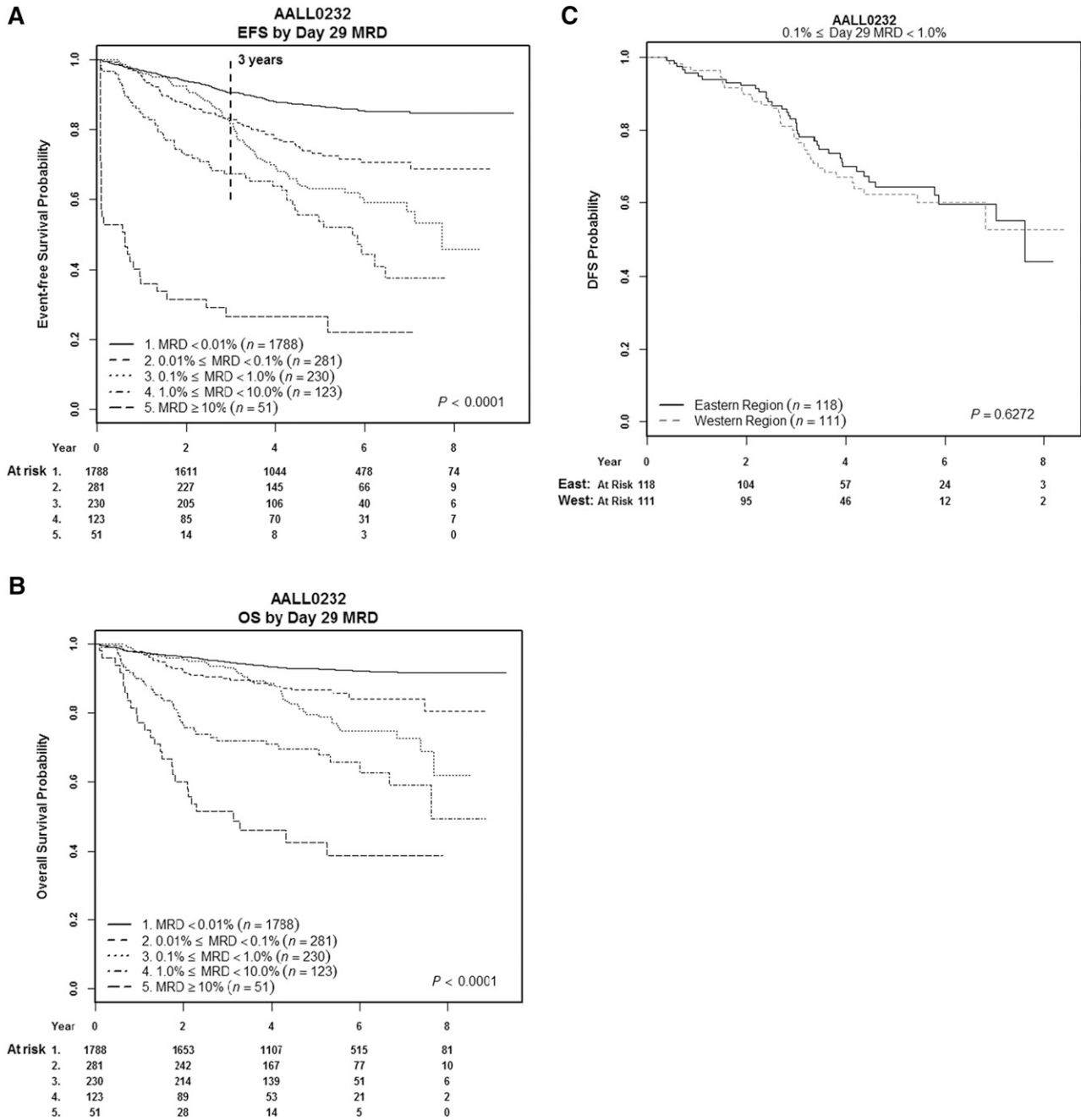
Figure 2. EFS of subjects with HR ALL treated on AALL0232 as a function of MRD levels at end of induction. (A) Five-year EFS for patients with <0.01% MRD was 87% ± 1%, 74% ± 4% for those between 0.01% and 0.1%, 63% ± 4% for those between 0.1% and 1%, 44% ± 9% for those between 1% and 10%, and 26% ± 9% for those >10%. The early outcome for subjects with 0.1% to 1% MRD is superior to that of subjects with lower levels of MRD, likely reflecting the effect of intensified therapy, though the curves cross at ~3 years (dotted line), and the overall outcome of this group is inferior to those with less MRD. (B) OS of subjects whose EFS was shown in panel A. The curves for OS in the cohort of subjects with 0.1% to 1% MRD again crosses that of those with lower MRD levels, but the time of the cross is delayed by ~1 year. (C) EFS of subjects with 0.1% to 1% MRD as a function of the location of their MRD testing. The unusual shape of this curve is identical in the 2 sites.

That this curve shape is not likely to be a statistical artifact is further illustrated by the fact that identical inflection points were seen if subjects studied at UW or JHU were compared (Figure 2C). This agreement also serves to indicate the reliability of the MRD assay in both laboratories.

Figure 2B shows the OS of patients as a function of MRD. As expected, MRD is also a prognostic factor for OS, with MRD-negative patients having a 93% ± 3% 8-year OS. Interestingly, the inflection point of the OS curve for patients with MRD between 0.1% and 1% is

shifted to the right by ~1 year, indicating that the therapeutic intensification for this group delayed but did not prevent relapse and death.

Subjects with an SER by BM morphology or MRD were eligible to have an additional MRD measurement performed at about week 12 of therapy, at the start of the first IM phase. These results were not provided to treating physicians and not used to make therapeutic interventions. As shown in Figure 3, persistence of MRD (≥0.01%) at this time point was associated with poor outcome: MRD-positive subjects
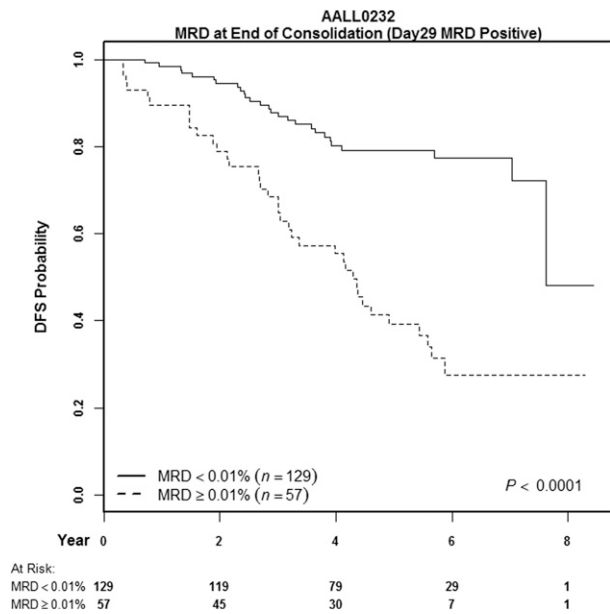
AALL0232
MRD at End of Consolidation (Day29 MRD Positive)

MRD < 0.01% (n = 129)
MRD ≥ 0.01% (n = 57)

P < 0.0001

At Risk:					
MRD < 0.01%	129	119	79	29	1
MRD ≥ 0.01%	57	45	30	7	1

Figure 3. DFS of patients with HR ALL as a function of MRD at end of consolidation. This figure displays DFS instead of EFS, because only those patients continuing on protocol therapy after induction (or extended induction) are included. Only subjects who were slow early responders at end induction had a second sample obtained, but this graph displays results only for those who were MRD positive at day 29. The 5-year DFS of subjects who cleared their MRD was quite good at 79% ± 5%, whereas those who were persistently positive had only a 39% ± 7% 5-year DFS.

had a 39% ± 7% 5-year DFS, compared with 79% ± 5% for those who converted from MRD positive to negative ($P < .0001$).

As noted above, sensitivity of detection of MRD was typically 1/10 000 cells. However, in ~10% of subjects in which the leukemic cell phenotype was markedly different from that seen in mature or immature B cells in the background, it was possible to identify small numbers of events in "empty space" that appeared to be leukemic. Because so few events were present, the confidence with which a call could be made was limited and quantitation not possible. These subjects with "suspicious" MRD are included in the MRD-negative curves in Figure 2, but when outcome of these subjects was compared with the remaining cohort of definitively negative subjects, they were shown to have a small but statistically significant inferior outcome (Figure 4), with a 5-year EFS of 81% ± 3% compared with 88% ± 1% ($P = .03$).

Relationship of MRD and methotrexate therapy

As described in more detail elsewhere, HD-MTX was superior to C-MTX as a postinduction regimen.[19] However, the magnitude of the therapy effect was small compared with that of MRD, which was highly prognostic in both treatment arms (Table 3). MRD-negative (<0.01%) subjects treated with C-MTX had an 86% ± 2% 5-year EFS, whereas MRD-positive subjects had only a 58% ± 4% 5-year EFS. For subjects treated with HD-MTX, the corresponding 5-year EFS rates were 88% ± 2% and 68% ± 4%, respectively.

Relationship of MRD and steroid therapy

MRD was prognostic in subjects receiving either Dex or Pred in induction: MRD-negative subjects treated with Dex had a 5-year EFS of 89% ± 2% compared with 65% ± 4% for those who were MRD positive ($P < .0001$); for subjects treated with Pred, MRD-negative subjects had a 5-year EFS of 86% ± 2%, whereas those who were MRD

positive had a 5-year EFS of 61% ± 4% ($P < .0001$). Although there was no overall difference between Dex and Pred, Dex was superior in a subset of subjects.[20] Specifically, subjects <10 years of age who were treated on the HD-MTX arm had a better outcome (88% ± 5% vs 81% ± 6%, $P = .05$) if they were treated with Dex. However, this difference in outcome could not have been predicted by looking at MRD following induction therapy. If anything, subjects younger than 10 who received the superior Dex regimen were more likely to be MRD positive (88/407, 21.6%) compared with those that received Pred (72/416, 17.3%).

Multivariate analysis of factors affecting MRD

Table 4 shows the results of a multivariate analysis using a Cox proportional hazards model including age, white blood cell count, MRD level, treatment arm, and day 15 morphologic assessment of SER. In this analysis, MRD had the highest hazard ratio (2.4) for predicting DFS. Note also that even though this was an NCI HR cohort, age and white blood cell count both independently predicted outcome. After adjusting for MRD, morphologic assessment of SER was not significant in the model.

Discussion

Numerous studies have established that MRD is the most powerful prognostic factor for predicting outcome in children, adolescents, and young adults with ALL.[1-6] Most clinical trials now use MRD levels as part of risk assignment, with those who are MRD positive assigned to a more intensive therapeutic regimen.[7,9-12,15] COG AALL0232 was designed to provide, in a nonrandomized fashion, more aggressive therapy for the SER patients. Although our previously completed study, COG P9900, showed that MRD levels between 0.01% and 0.1% were associated with poor outcome in HR subjects,[1] results of that study were not available at the time of the design of AALL0232, and one study using flow cytometry for MRD detection in patients treated on a Berlin-Frankfurt-Munster backbone had previously shown 0.1% to be a prognostic cutpoint.[5]

MRD >0.01% at end of induction was again found to be prognostic in this study. Persistence of MRD at 12 weeks was found to be a very poor prognostic factor, with 5-year DFS of ~40%, but patients who were MRD-positive at end induction but became MRD <0.01% by then did very well, with a nearly 80% 5-year DFS. In contrast to other studies that use end-consolidation MRD >0.1% as a cutoff for risk stratification,[7] our data suggested that 0.01% was a better discriminator; patients who were MRD negative at the 0.1% cutoff had only a 70% ± 7% 5-year DFS (not shown). However, the MRD-positive group in Figure 3 mainly included patients with end-induction MRD >0.1% plus only 8 patients with MRD between 0.01% and 0.1% who were also morphologic slow early responders. Thus, we do not believe we can conclude that patients with levels of MRD between 0.01% and 0.1% at both time points have a very bad outcome.

Although HD-MTX was associated with superior outcome overall and for both rapid early response and SER subjects, the differences in outcome between HD-MTX and C-MTX arm were only a few percentage points, whereas outcome was 20 (for HD-MTX) and 28 (for C-MTX) percentage points worse for MRD-positive patients compared with those who were negative. Thus, although better chemotherapy can improve outcome, the effect of MRD remains dramatic.

In this trial, we attempted to improve the outcome of subjects who were end-induction MRD positive ≥0.1% by providing ~4 extra
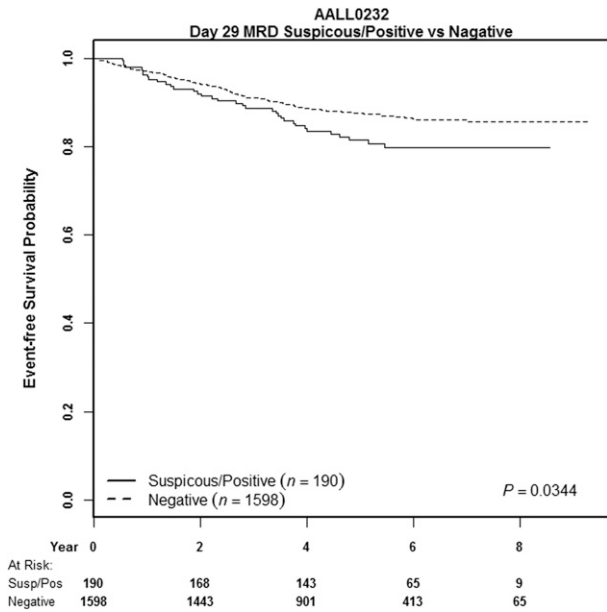
Figure 4. Prognostic significance of very low levels of MRD. In ~10% of cases that were considered MRD negative (at 0.01%), a few events with an abnormal phentotype were detected, though quantitation of such a small population is not accurate. The EFS of such subjects was slightly poorer than that of those who were completely negative by flow cytometry (5-year EFS, 81% ± 3% vs 88% ± 1%).

months of intensive chemotherapy. Ultimately, the 5-year EFS of the SER subjects was intermediate between those with 0.01% to <0.1% MRD and those with ≥1% MRD. However, the shape of the EFS and OS curves of this cohort was unusual, with early relapse and death rates lower than that seen with subjects with lower levels of MRD and similar to that of MRD-negative subjects. This suggests to us that the strategy of trying to intensify therapy had some clinical effect; although it might have been more successful in patients with lower levels of MRD, the short course of intensive therapy was insufficient to overcome the adverse effect of higher levels of MRD.

It is interesting to speculate on the possible implications of the lower rate of early relapse in the cohort of subjects with MRD between 0.1% and 1%. The biology of early relapse is known to be different from that of late relapse,[21] and the outcome of the latter group is better.[22] The OS curve for this cohort showed a time-shifted but maintained inflection point, suggesting that patients whose relapse was delayed still had biological characteristics of "early relapse." It would be interesting to compare blasts from patients relapsing after 3 years in the 0.01% to 0.1% and 0.1% to 1% cohorts to see if they differed.

There are limited data that speak directly to the question of whether the poor prognostic effect of MRD can be overcome with therapy. The COG showed previously that intensification of postinduction therapy improved EFS and OS for subjects with a poor morphologic early response,[23] suggesting that a similar strategy might be successful when response was assessed with MRD. Answering this question directly, however, has been difficult, because rather than randomizing MRD-positive subjects to receive more intensive therapy or not, most studies have used assessment of MRD to change risk assignment. However, The United Kingdom ALL2003 trial, using a treatment backbone similar to that used in AALL0232, recently randomized subjects with day 29 MRD ≥0.01% who were otherwise classified as good or intermediate risk based on clinical features to receive more or less intensive postinduction therapy.[10] In that study, the 5-year EFS of those with MRD ≥0.01% randomized to augmented therapy was superior to those receiving standard therapy (89.6% [95% confidence interval, 85.9-93.5]

vs 82.8% [78.1-87.5]; P = .04). Also, in both pediatric[24] and adult ALL, it has been shown that MRD-positive subjects who receive marrow transplantation fare better than those who do not,[25] suggesting that the poor prognosis of some MRD-positive subjects can be overcome by specific therapeutic interventions.

Most studies of MRD in pediatric ALL rely on molecular determination of MRD, using antigen-receptor PCR studies to detect clone-specific markers. This technique has been well standardized, and with it, MRD can be determined reproducibly in many laboratories.[26-31] There are fewer data on reproducibility of MRD by flow cytometry, with one study indicating that reproducibility is good down to levels of 0.1% but not as good at 0.01%.[32] In our prior study,[1] all MRD results were obtained in a single laboratory, but the size of this study made it necessary to involve a second reference laboratory. Our results, using a standardized protocol, indicate that 2 experienced laboratories can get excellent agreement as evidenced by the frequency with which MRD positivity is detected at any threshold level cutoff, comparison of outcome data of subjects tested in both labs, and limited sample exchange. This indicates that flow cytometry, which has the advantages of being faster and less expensive than antigen-receptor PCR, can be used for MRD assessment in the context of a clinical trial with MRD determinations performed at multiple sites.

Our studies used 6-color flow cytometry with a relatively limited panel designed to be able to efficiently process the very large numbers of subjects enrolled on this trial in only 2 labs. Although 8- to 10-color flow cytometry methods are now becoming standard, our studies began before these were widely available for clinical use. However, as other studies have shown,[32-34] more extensive training and validation studies would be necessary before this technique could be routinely employed by a larger number of laboratories. We have now implemented a process to significantly expand the number of laboratories that can perform MRD in the context of COG ALL clinical trials.

There has been significant recent interest in the potential use of MRD as a surrogate marker to help to evaluate the efficacy of new therapeutic agents.[35-39] Given the number of potential new agents, new strategies should be considered that do not rely on long-term EFS to assess whether or not an agent is effective, particularly in pediatric ALL, where numbers of patients eligible for new agent trials is small. It is very attractive to use MRD response as an outcome measure to help to select effective agents. To do this, however, requires evidence to show that changes in MRD levels in patients receiving a new agent ultimately mirror changes in outcome. Results from this study provide a caution to the wholesale use of MRD for this purpose. Specifically, subjects <10 years who were treated with HD-MTX had a superior outcome when they were given Dex in induction compared with Pred. However, this was not associated with an improvement of end-induction MRD rates. Thus, had end-induction MRD been a surrogate marker used to assess this intervention, Dex would have been erroneously concluded to be ineffective. This negative result could, however, possibly be explained because subjects randomized to Dex received this agent only on days 1 to 14 and thus had no steroid therapy at all for 2 weeks

Table 3. Prognostic significance of MRD in methotrexate cohorts and of HD-MTX in MRD-positive and negative subjects

End-induction MRD	C-MTX 5-y EFS	HD-MTX 5-y EFS	P value
<0.01%	86% ± 2%	88% ± 2%	.19
≥0.01%	58% ± 4%	68% ± 4%	.008
P value	<.0001	<.0001	—

Table 4. Multivariate analyses

Parameter	χ^2	Probability $> \chi^2$	Hazard ratio	95% hazard ratio CI	
MRD day 29 ≥0.01%	70.598	<.0001	2.433	1.978	2.994
Age ≥10 y	35.433	<.0001	2.341	1.769	3.097
WBC ≥50 × 10⁹/L	37.605	<.0001	2.162	1.690	2.766
Treatment: Capizzi	6.125	.013	1.277	1.052	1.549
Day 15 morphology: SER	1.536	.215	1.194	0.902	1.579

WBC, white blood cell.

prior to the time of MRD assessment at day 29, whereas those randomized to Pred had no steroid-free gap prior to determining MRD at day 29. A similar caution arose from a UK ALL trial that showed a positive effect of mitoxantrone compared with idarubicin in induction of children and adolescents with first relapse of ALL. In spite of the significantly superior outcome of those randomized to receive mitoxantrone, there was no difference in end induction MRD levels between the mitoxantrone and idarubicin arms,[40] though it should be noted that in that study, MRD results were only available on a selected subset of subjects and the number of subjects analyzed was very small. Collectively, these results highlight the need to carefully design any study in which MRD might be proposed as a surrogate marker.

In summary, these results show the continued major importance of MRD analysis for risk assignment of subjects with high-risk B-ALL, even as therapies for this disease improve overall outcome. The best MRD threshold to identify good-risk groups at day 29 of induction therapy was 0.01%. The results further show that multiparameter flow cytometry is an effective method for measuring MRD and that the procedure can be standardized between 2 experienced laboratories to provide equivalent results that ensure subjects receive appropriate therapy no matter where their treatment is given.

Acknowledgments

Authorship

Contribution: M.J.B. designed research, performed research, analyzed data, and wrote the paper; B.L.W. performed research, contributed analytical tools, analyzed data, and wrote the paper; M.D. designed research, analyzed data, and wrote the paper; M.L.L. and E.A.R. performed research, analyzed data, and wrote the paper; W.L.S. designed research, performed research, analyzed data, and wrote the paper; J.B.N. designed and performed research; A.J.C., N.A.H., J.M.G.-F., and C.L.W. performed research; Y.D. analyzed data and wrote the paper; and N.J.W., S.P.H., W.L.C., and E.L. designed and performed research, analyzed data, and wrote the paper.

Conflict-of-interest disclosure: The authors declare no competing financial interests.

James B. Nachman died on June 10, 2011.

Correspondence: Michael J. Borowitz, Johns Hopkins Medical Institutions, 401 N Broadway, Weinberg 2335, Baltimore, MD 21287; e-mail: mborowit@jhmi.edu.

References

1. Borowitz MJ, Devidas M, Hunger SP, et al; Children's Oncology Group. 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):5477-5485.

2. Cazzaniga G, Gaipa G, Rossi V, Biondi A. Minimal residual disease as a surrogate marker for risk assignment to ALL patients. *Rev Clin Exp Hematol.* 2003;7(3):292-323.

3. Conter V, Aricò M, Basso G, et al; Associazione Italiana di Ematologia ed Oncologia Pediatrica. Long-term results of the Italian Association of Pediatric Hematology and Oncology (AIEOP) Studies 82, 87, 88, 91 and 95 for childhood acute lymphoblastic leukemia. *Leukemia.* 2010;24(2):255-264.

4. Coustan-Smith E, Sancho J, Hancock ML, et al. Clinical importance of minimal residual disease in childhood acute lymphoblastic leukemia. *Blood.* 2000;96(8):2691-2696.

5. Dworzak MN, Fröschl G, Printz D, et al; Austrian Berlin-Frankfurt-Münster Study Group. Prognostic significance and modalities of flow cytometric minimal residual disease detection in childhood acute lymphoblastic leukemia. *Blood.* 2002;99(6):1952-1958.

6. Panzer-Grümayer ER, Schneider M, Panzer S, Fasching K, Gadner H. Rapid molecular response during early induction chemotherapy predicts a good outcome in childhood acute lymphoblastic leukemia. *Blood.* 2000;95(3):790-794.

7. Conter V, Bartram CR, Valsecchi MG, 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):3206-3214.

8. Coustan-Smith E, Sancho J, Behm FG, et al. Prognostic importance of measuring early clearance of leukemic cells by flow cytometry in childhood acute lymphoblastic leukemia. *Blood.* 2002;100(1):52-58.

9. Pui CH, Campana D, Pei D, et al. Treating childhood acute lymphoblastic leukemia without cranial irradiation. *N Engl J Med.* 2009;360(26):2730-2741.

10. Vora A, Goulden N, Mitchell 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):809-818.

11. Yeoh AE, Ariffin H, Chai EL, et al. Minimal residual disease-guided treatment deintensification for children with acute lymphoblastic leukemia: results from the Malaysia-Singapore acute lymphoblastic leukemia 2003 study. *J Clin Oncol.* 2012;30(19):2384-2392.

12. Yamaji K, Okamoto T, Yokota S, et al; Japanese Childhood Cancer Leukemia Study Group. Minimal residual disease-based augmented therapy in childhood acute lymphoblastic leukemia: a report from the Japanese Childhood Cancer and Leukemia Study Group. *Pediatr Blood Cancer.* 2010;55(7):1287-1295.

13. Escherich G, Horstmann MA, Zimmermann M, Janka-Schaub GE; COALL study group. Cooperative study group for childhood acute lymphoblastic leukaemia (COALL): long-term results of trials 82,85,89,92 and 97. *Leukemia.* 2010;24(2):298-308.

14. Schmiegelow K, Forestier E, Hellebostad M, et al; Nordic Society of Paediatric Haematology and Oncology. Long-term results of NOPHO ALL-92 and ALL-2000 studies of childhood acute lymphoblastic leukemia. *Leukemia.* 2010;24(2):345-354.

15. Marshall GM, Dalla Pozza L, Sutton R, et al. High-risk childhood acute lymphoblastic leukaemia in first remission treated with novel intensive chemotherapy and allogeneic transplantation. *Leukemia.* 2013;27(7):1497-1503.

16. Weir EG, Cowan K, LeBeau P, Borowitz MJ. A limited antibody panel can distinguish B-precursor acute lymphoblastic leukemia from normal B precursors with four color flow cytometry: implications for residual disease detection. *Leukemia.* 1999;13(4):558-567.

17. Dworzak MN, Fritsch G, Panzer-Grümayer ER, Mann G, Gadner H. Detection of residual disease in pediatric B-cell precursor acute lymphoblastic leukemia by comparative phenotype mapping: method and significance. *Leuk Lymphoma.* 2000;38(3-4):295-308.

18. Borowitz MJ, Pullen DJ, Shuster JJ, et al; Children's Oncology Group study. Minimal residual disease detection in childhood precursor-B-cell acute lymphoblastic leukemia: relation to other risk factors. A Children's Oncology Group study. *Leukemia.* 2003;17(8):1566-1572.

19. Larsen EC, Salzer WL, Devidas M, et al. Comparison of high-dose methotrexate (HD-MTX) with Capizzi methotrexate plus asparaginase (C-MTX/ASNase) in children and young adults with high-risk acute lymphoblastic leukemia (HR-ALL): a report from the Children's Oncology Group Study AALL0232 [abstract]. *J Clin Oncol.* 2011;29(suppl 15). Abstract 3.

20. Winick NL, Salzer WL, Devidas M, et al. Dexamethasone (DEX) versus prednisone (PRED) during induction for children with high-risk acute lymphoblastic leukemia (HR-ALL): a report from the Children's Oncology Group Study AALL0232. *J Clin Oncol.* 2011;29(suppl 15). Abstract 9504.

21. Bhojwani D, Kang H, Moskowitz NP, et al. Biologic pathways associated with relapse in childhood acute lymphoblastic leukemia: a Children's Oncology Group study. *Blood.* 2006; 108(2):711-717.

22. Raetz EA, Borowitz MJ, Devidas M, 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): 3971-3978.

23. Nachman JB, Sather HN, Sensel MG, et al. Augmented post-induction therapy for children with high-risk acute lymphoblastic leukemia and a slow response to initial therapy. *N Engl J Med.* 1998;338(23):1663-1671.

24. Eckert C, Henze G, Seeger K, 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):2736-2742.

25. Gökbuget N, Stanze D, Beck J, et al; German Multicenter Study Group for Adult Acute Lymphoblastic Leukemia. Outcome of relapsed adult lymphoblastic leukemia depends on response to salvage chemotherapy, prognostic factors, and performance of stem cell transplantation. *Blood.* 2012;120(10):2032-2041.

26. Gabert J, Beillard E, van der Velden VH, 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(12):2318-2357.

27. Pongers-Willemse MJ, Seriu T, Stolz F, et al. Primers and protocols for standardized detection of minimal residual disease in acute lymphoblastic leukemia using immunoglobulin and T cell receptor gene rearrangements and TAL1 deletions as PCR targets: report of the BIOMED-1 CONCERTED ACTION: investigation of minimal residual disease in acute leukemia. *Leukemia.* 1999;13(1):110-118.

28. Salari F, Shahjahani M, Shahrabi S, Saki N. Minimal residual disease in acute lymphoblastic leukemia: optimal methods and clinical relevance, pitfalls and recent approaches. *Med Oncol.* 2014; 31(11):266.

29. van der Velden VH, Hochhaus A, Cazzaniga G, Szczepanski T, Gabert J, van Dongen JJ. Detection of minimal residual disease in hematologic malignancies by real-time quantitative PCR: principles, approaches, and laboratory aspects. *Leukemia.* 2003;17(6): 1013-1034.

30. van der Velden VH, Panzer-Grümayer ER, Cazzaniga G, et al. Optimization of PCR-based minimal residual disease diagnostics for childhood acute lymphoblastic leukemia in a multi-center setting. *Leukemia.* 2007;21(4):706-713.

31. Brüggemann M, Schrauder A, Raff T, et al; European Working Group for Adult Acute Lymphoblastic Leukemia (EWALL); International Berlin-Frankfurt-Münster Study Group (I-BFM-SG). 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):521-535.

32. Dworzak MN, Gaipa G, Ratei R, et al. Standardization of flow cytometric minimal residual disease evaluation in acute lymphoblastic leukemia: Multicentric assessment is feasible. *Cytometry B Clin Cytom.* 2008;74(6):331-340.

33. Björklund E, Matinlauri I, Tierens A, et al. Quality control of flow cytometry data analysis for evaluation of minimal residual disease in bone marrow from acute leukemia patients during treatment. *J Pediatr Hematol Oncol.* 2009;31(6): 406-415.

34. Fossat C, Roussel M, Arnoux I, et al; French Multicenter Study Groups for Pediatric and Adult ALL. Methodological aspects of minimal residual disease assessment by flow cytometry in acute lymphoblastic leukemia: a French multicenter study. *Cytometry B Clin Cytom.* 2015;88(1): 21-29.

35. Pui CH, Mullighan CG, Evans WE, Relling MV. Pediatric acute lymphoblastic leukemia: where are we going and how do we get there? *Blood.* 2012;120(6):1165-1174.

36. Topp MS, Kufer P, Gökbuget N, et al. Targeted therapy with the T-cell-engaging antibody blinatumomab of chemotherapy-refractory minimal residual disease in B-lineage acute lymphoblastic leukemia patients results in high response rate and prolonged leukemia-free survival. *J Clin Oncol.* 2011;29(18):2493-2498.

37. Chevallier P, Bodet-Milin C, Robillard N, et al. BCR-ABL1 molecular remission after 90Y-epratuzumab tetraxetan radioimmunotherapy in CD22+ Ph+ B-ALL: proof of principle. *Eur J Haematol.* 2013;91(6):552-556.

38. Brentjens RJ, Davila ML, Riviere I, et al. CD19-targeted T cells rapidly induce molecular remissions in adults with chemotherapy-refractory acute lymphoblastic leukemia. *Sci Transl Med.* 2013;5(177):177ra38.

39. Schultz KR, Bowman WP, Aledo A, et al. Improved early event-free survival with imatinib in Philadelphia chromosome-positive acute lymphoblastic leukemia: a children's oncology group study. *J Clin Oncol.* 2009;27(31): 5175-5181.

40. Parker C, Waters R, Leighton C, et al. Effect of mitoxantrone on outcome of children with first relapse of acute lymphoblastic leukaemia (ALL R3): an open-label randomised trial. *Lancet.* 2010; 376(9757):2009-2017.