Time point-dependent concordance of flow cytometry and real-time quantitative polymerase chain reaction for minimal residual disease detection in childhood acute lymphoblastic leukemia

Giuseppe Gaipa,[1] Giovanni Cazzaniga,[1] Maria Grazia Valsecchi,[2] Renate Panzer-Grümayer,[3] Barbara Buldini,[4] Daniela Silvestri,[2] Leonid Karawajew,[5] Oscar Maglia,[1] Richard Ratei,[5] Alessandra Benetello,[4] Simona Sala,[1] Angela Schumich,[3] Andre Schrauder,[6] Tiziana Villa,[1] Marinella Veltroni,[7] Wolf-Dieter Ludwig,[5] Valentino Conter,[8] Martin Schrappe,[6] Andrea Biondi,[1] Michael N. Dworzak,[3] and Giuseppe Basso[4]

[1]M. Tettamanti Research Center, Pediatric Clinic University of Milano Bicocca, Monza, Italy; [2]Medical Statistics Unit, Department of Clinical and Preventive Medicine, University of Milano Bicocca, Monza, Italy; [3]Children's Cancer Research Institute and St. Anna Children's Hospital, Vienna, Austria; [4]Laboratorio di Oncoematologia Pediatrica, Department of Pediatrics, University of Padova, Padova Italy; [5]Hematology, Oncology and Tumor Immunology, Robert-Roessle-Clinic at the HELIOS Klinikum Berlin, Charité Medical School, Berlin, Germany; [6]Department of Pediatrics, University Medical Center Schleswig-Holstein, Campus Kiel, Kiel, Germany; [7]Department of Pediatric Hematology Oncology, A.O.U. Meyer, Firenze, Italy, and [8]Department of Pediatrics, Ospedali Riuniti di Bergamo, Bergamo, Italy

ABSTRACT

Background

Flow cytometric analysis of leukemia-associated immunophenotypes and polymerase chain reaction-based amplification of antigen-receptor genes rearrangements are reliable methods for monitoring minimal residual disease. The aim of this study was to compare the performances of these two methodologies in the detection of minimal residual disease in childhood acute lymphoblastic leukemia.

Design and Methods

Polymerase chain reaction and flow cytometry were simultaneously applied for prospective minimal residual disease measurements at days 15, 33 and 78 of induction therapy on 3565 samples from 1547 children with acute lymphoblastic leukemia enrolled into the AIEOP-BFM ALL 2000 trial.

Results

The overall concordance was 80%, but different results were observed according to the time point. Most discordances were found at day 33 (concordance rate 70%) in samples that had significantly lower minimal residual disease. However, the discordance was not due to different starting materials (total *versus* mononucleated cells), but rather to cell input number. At day 33, cases with minimal residual disease below or above the 0.01% cut-off by both methods showed a very good outcome (5-year event-free survival, 91.6%) or a poor one (5-year event-free survival, 50.9%), respectively, whereas discordant cases showed similar event-free survival rates (around 80%).

Conclusions

Within the current BFM-based protocols, flow cytometry and polymerase chain reaction cannot simply substitute each other at single time points, and the concordance rates between their results depend largely on the time at which they are used. Our findings suggest a potential complementary role of the two technologies in optimizing risk stratification in future clinical trials.

Key words: childhood ALL, minimal residual disease, prognostic value.

Citation: Gaipa G, Cazzaniga G, Valsecchi MG, Panzer-Grümayer R, Buldini B, Silvestri D, Karawajew L, Maglia O, Ratei R, Benetello A, Sala S, Schumich A, Schrauder A, Villa T, Veltroni M, Ludwig W-D, Conter V, Schrappe M, Biondi A, Dworzak MN, and Basso G. Time point-dependent concordance of flow cytometry and real-time quantitative polymerase chain reaction for minimal residual disease detection in childhood acute lymphoblastic leukemia. Haematologica 2012;97(10):1586-1593. doi:10.3324/haematol.2011.060426

©2012 Ferrata Storti Foundation. This is an open-access paper.

Acknowledgments: we thank the partners in the AIEOP and BFM reference laboratories and in the central PCR- and FCM-MRD laboratories, all the technicians for their expert work and the data managers for their careful study conduct.

Funding: for AIEOP: this work was supported by Comitato M.L. Verga and Fondazione Tettamanti, Fondazione Città della Speranza, Fondazione Cariparo and Grant Ric. Corrente OBG 2006/02/R/001822, Associazione Italiana per la Ricerca sul Cancro (AIRC; to ABi, GC, MGV), Fondazione Cariplo (ABi), Ministero dell'Istruzione, Università e Ricerca (MIUR; ABi), Associazione Gian Franco Lupo. We thank AIEOP reference laboratories (Padova and Monza) and Prof. G. Masera for his thoughtful supervision of the study. For BFM: this work was supported by the Austrian National Bank; Grant number: 10962 (to MND), by the Österreichische Kinderkrebsforschung to R Panzer-Grümayer (BFM-AT reference laboratory:), by Deutsche Krebshilfe, Bonn, Germany (grant n. 50-2698 Schr1 and grant n. 50-2410 Ba7), by the Competence Network Pediatric Hematology and Oncology (KPOH) which was funded by the Federal Ministry of Research (BMBF) and Oncosuisse/Krebsforschung Schweiz (Grant OCS 1230-02-2002). We thank Prof. G. Mann for his support of this study.

Manuscript received on December 11, 2011. Revised version arrived on April 19, 2012. Manuscript accepted on April 19, 2012.

Correspondence:
Andrea Biondi, MD, M. Tettamanti Research Center, University of Milano Bicocca, via Pergolesi, 33 20052 Monza, MI.
E-mail: abiondi.unimib@gmail.com/ andrea.biondi@unimib.it

The online version of this article has a Supplementary Appendix.

Introduction

Minimal residual disease (MRD) is a powerful predictor of response to treatment and clinical outcome in childhood acute lymphoblastic leukemia (ALL).[1-10] The most reliable methods for studying MRD are the flow cytometric (FCM) analysis of leukemia-associated immunophenotypes and polymerase chain reaction (PCR) amplification of antigen-receptor rearrangements.[11-19]

FCM is more time- and cost-effective, whereas PCR has a higher sensitivity. Wide standardization, quality control and guidelines on real time quantitative PCR analysis and interpretation have been assessed within the "EuroMRD' group (previous 'European Study Group for MRD detection in ALL", ESG-MRD ALL).[16] A quantifiable range achievable by FCM-MRD can generally be defined as the ability to detect 30 clustered MRD events in 3×10^5 total cellular events (0.01%). However, a cluster of at least 10 events with leukemia-associated immunophenotype and back-gating light scatter can be sufficient to define a sample as "MRD-positive".[19]

Several parameters are critical for the conception and interpretation of MRD studies, including therapeutic context, selection of patients, sample timing, target markers, the required sensitivity of the assay, inter-laboratory standardization (particularly relevant in multicenter studies), and the retrospective or prospective design of the study. Another critical issue is the discordance between PCR and FCM results. This may be mainly due to the limited number of total cells available in the FCM test-tube, which reduces the sensitivity, and thus the accuracy, of FCM at low levels of MRD (below 10^{-4}). However, qualitative or quantitative discrepancies can also be explained by different factors affecting either PCR or FCM. These factors include: (i) quality of clonal PCR-markers; (ii) non-specific amplification of normal DNA or amplification of DNA from dead/dying leukemia cells; (iii) oligoclonality and/or clonal evolution, (iv) age-related[20] or therapy-related bone marrow B-cell precursor regeneration status;[21,22] (v) immunophenotypic changes between diagnosis and relapse;[23,24] and (vi) drug-induced immunophenotypic modulation.[25]

PCR-MRD was largely used to stratify patients within the AIEOP-BFM ALL 2000 protocol.[9,18] However, few studies have compared the MRD estimates obtained by the simultaneous application of FCM and PCR in therapeutic protocols and at specific time points.[26-30] Indeed, PCR and FCM MRD measurements may be combined for a better definition of the risk of treatment failure in children with ALL treated in a multicenter study.[31]

Here we report the results of the contemporary application of FCM and PCR analysis to 3565 bone marrow samples. The samples were obtained from 1547 consecutive children with ALL at multiple time points of the induction phase within the AIEOP-BFM ALL 2000 trial. The impact of different cell preparation procedures and different FCM methodological options was also tested. This is the largest reported cooperative study that has simultaneously evaluated the performances of FCM and PCR in measuring MRD in children with ALL.

Design and Methods

Study population

Between September 2000 and June 2006, 4741 children aged 1 to <18 years with newly diagnosed Philadelphia chromosome-negative ALL were consecutively enrolled in the AIEOP-BFM ALL 2000 study in Italy (NCT00613457), Austria, Germany and Switzerland (NCT00430118). Out of these, 3715 patients (78%) were stratified for treatment according to MRD data measured in bone marrow by PCR.[9]

In a subset of 1547 of the 4741 children (32.6%), MRD was also measured by four-color FCM on days 33 and 78, and in a subset of 471 patients MRD was measured in bone marrow on day 15. The sensitivity of both FCM and PCR was at least 1×10^{-4}. Continuous PCR log values suitable for quantitative comparison with FCM results were available from 1115 patients, while in 432 out of 1547 patients PCR-MRD levels were quantified only in terms of log range. For example, an MRD value falling into the range $\geq5\times10^{-5}$ and $<5\times10^{-4}$ was entered as '10^{-4}', but the actual continuous MRD values were unspecified. These samples were used only for qualitative comparison with FCM (positive *versus* negative), but not for continuous assessment of MRD level. FCM and PCR data were managed independently. A separate series of 110 patients consecutively recruited in the MRD-laboratory in Vienna were analyzed for a triple comparison study with both four- and seven-color FCM.

Since PCR-MRD analysis of samples on days 33 and 78 had therapeutic implications in the AIEOP-BFM ALL 2000 protocol, priority was given to this technique. It was applicable in 78% of patients,[18] while FCM-MRD was performed only when excess material was available; the assessment of overall FCM-MRD applicability was not, therefore, possible. However, virtually all patients studied (>98%) showed suitable immunological markers for FCM-MRD monitoring at diagnosis.

Institutional ethical committees approved the bone marrow sampling for diagnostic work-up and MRD investigation along with the international trial. Each was performed strictly according to informed consent guidelines.

Diagnostic studies

Bone marrow samples were tested by references laboratories in each country based on standard analyses. Immunophenotyping was performed by multiparametric FCM, as reported elsewhere.[32] Leukemia-associated immunophenotypes were assessed according to a previously reported method.[19,25]

Flow cytometric minimal residual disease study

For each follow-up specimen, the following procedures were carried out: one aliquot of total nucleated cells (NC) was used for the FCM assay, while one aliquot underwent Ficoll gradient separation and the mononucleated cell (MNC) fraction was used for the PCR assay. The following monoclonal antibodies were used in four-color combinations for the detection of MRD:[19] TdT (Supertechs, Bethesda, MD, USA); CD3, CD5, CD10, CD11a, CD20, CD34, CD38, CD45 (Becton Dickinson, San Jose, CA, USA); and CD10 (PE), CD58, CD19, CD99, CD7, CD3 (PE-Cy7) (Beckman-Coulter, Miami, FL, USA). Monoclonal antibodies were combined as indicated in *Online Supplementary Table S1*. Quadruple combinations were defined based on those markers that proved to be most relevant for MRD detection in ALL.[5,20,21,25,33-35] For B-cell precursor-ALL, combinations included triple-marker back bones useful for tracking leukemia-associated phenotypes in different tubes, such as CD10/CD19/CD34 and CD10/CD19/CD45. For T-ALL, however, CD5 or CD7 in combination with CD3 was mainly used to ensure blast identification in addition to aberrant or ectopic expression of CD99 and TdT, respectively.[36]

For FCM-MRD measurements, at least 3×10^5 ungated events were collected and analyzed.[19] The minimum target sensitivity for quantifying MRD was defined as the ability to detect 30 clustered

MRD events in 3×10^5 total cellular events (0.01%). However, a cluster of at least ten events with leukemia-associated immunophenotype and back-gating light scatter[19] was sufficient to define a sample as "MRD-positive".

FCM analyses were performed in four laboratories, Berlin, Monza, Padova, and Vienna, which have shared validated inter-laboratory standardization procedures since 2000 for the purposes of the AIEOP-BFM ALL trials (recruitment is approximately 1000 pediatric ALL patients per year). In particular, inter-center reproducibility was assessed in several steps according to standardized procedures.[19,25,36]

Triple-comparisons (polymerase chain reaction versus flow cytometry on mononucleated cells or total nucleated cells)

In order to investigate the variables related to sample preparation and analysis more thoroughly, we investigated 266 follow-up samples (day 15: n=55, day 33: n=69, day 52: n=55, and day 78: n=87) from 110 patients (96 with B-cell precursor-ALL and 14 with T-ALL) consecutively recruited in the Vienna laboratory. After preparation of NC and MNC aliquots (see above), each MNC preparation was then further subdivided for PCR and FCM. The PCR assay was performed as described previously.[17] FCM analysis of MNC preparations (FCM^{MNC}) was done using a seven-color panel, analyzing 5×10^5 cells per tube, while FCM analysis of NC (FCM^{NC}) was carried out by four-color FCM in which 3×10^5 cells per tube were analyzed.[19] Exactly the same gating strategy was adopted for both FCM^{NC} and FCM^{MNC}. FCM data were interpreted according to the published method.[19]

Identification of molecular markers and minimal residual disease analysis

The logistics of the study, cell sample isolation and identification of the markers for MRD evaluation have been recently reported.[16,18] Briefly, DNA samples obtained at diagnosis from ALL cases were screened for *IGH* and *TCR* rearrangements.

Allele-specific oligonucleotide (ASO) primers were designed to complement the junctional region sequence of each target. MRD PCR targets were tested in order to select two for each patient. They each had to be sensitive to at least 10^{-4} and the quantitative range of the first target had to be at least 10^{-4}, while that of the second target had to be at least 5×10^{-4} .[17,18] Of note, 179 out of 1547 patients (11.6%) had one marker with a quantitative range of at least 10^{-4} and another with a quantitative range of 5×10^{-4}. All remaining patients had both markers with a quantitative range of at least 10^{-4}. For comparison, we used the highest MRD result obtained by either the two PCR targets or the FCM antibody combinations. Real-time quantitative PCR analysis was performed and interpreted according to the guidelines developed by the "European Study Group for MRD detection in ALL" (EuroMRD-ALL).[18]

Treatment

AIEOP-BFM ALL 2000 treatment phases have been recently described in detail.[9,31] Briefly, after 7 days of monotherapy with prednisone and one dose of intrathecal methotrexate on day 1, treatment was complemented with corticosteroids. This treatment phase was followed by intravenous cyclophosphamide and cytarabine, intrathecal methotrexate, and oral 6-mercaptopurine. Subsequent treatment elements differed by risk group.[9,18] Very high-risk patients were eligible for allogeneic stem cell transplantation.

Statistical analysis

Positive and negative predictive values, concordance, sensitivity and specificity were calculated to describe the two methods, taking PCR as the reference, with results expressed as continuous values. The two methods were compared using the Bland-Altman algorithm after logarithmic transformation.[37] Thus, the differences between the two measures on each subject were plotted against their average value. After excluding any dependence, the hypothesis of zero bias was examined with a paired t-test. The 95% range was calculated from twice the standard deviation. Box plots (horizontal lines at the 25th, 50th and 75th percentiles) and scatter plots were used to describe continuous values and Wilcoxon's test was used to compare medians. Event-free survival (EFS) was calculated as time from diagnosis to the first event i.e., resistance, relapse, death in complete clinical remission and second malignant neoplasm, and then compared with the log-rank test.

Results

Concordance between flow cytometry and polymerase chain reaction detection of minimal residual disease

The two methods were first compared for their ability to detect the presence of any residual leukemic cells. Overall, the concordance rate was 71%: discordant FCM-negative and PCR-positive samples were more prevalent (n=939/3565, 26.3%), whereas there were only 102 (2.9%) FCM-positive and PCR-negative samples. The concordance rate was 86% at day 15 (471 samples, range by center: 84%-93%), 60% at day 33 (1547 samples, range by center: 53%-69%) and 77% at day 78 (1547 samples, range by center: 75%-85%).

For a more reliable comparison between FCM and PCR, a precise cut-off of 0.01% was established for both methods and applied in the 1115 children for whom continuous values on PCR were available.

In the following, we will refer to a result ≥ 0.01% as MRD-positive and a result of <0.01% as MRD-negative. Overall, the two methods gave concordant results, either above or below the cut-off (0.01%), in 2160 of 2701 samples analyzed at the three time points (concordance rate 80%) (Table 1). Concordance was higher, however, at day 15 and day 78 (86% and 87%, respectively) than at day 33 (70%). The difference in overall concordance rate between each center was always <10%. Concordance was also evaluated separately by immunophenotype (B-cell precursor ALL *versus* T-ALL) with the same time trend in the two groups, but with a lower concordance rate in T-ALL (*Online Supplementary Table S2*).

The direct comparison of MRD estimates by the two techniques at each time point is shown in Figure 1 (panels A, B, C). For patients with positive MRD by both techniques, we examined the differences between the two methods more closely using the Bland-Altman algorithm (Figure 1, panels D, E, F). After logarithmic transformation, the estimated mean differences of FCM results *versus* PCR results were significantly different from zero, and were markedly consistent according to different levels of MRD.

At each time point, MRD discordant samples (FCM-positive but PCR-negative) tended to have a significantly lower positivity level by PCR (10^{-4} log-range) compared to concordant samples ($P<0.001$, *Online Supplementary Figure S1*). Analogously, among FCM-positive samples, discordant samples (PCR-negative) had significantly lower FCM MRD levels at each time point than did concordant samples ($P<0.001$, *Online Supplementary Figure S1*).

In samples that were positive for MRD as detected by

PCR, we also assessed whether the signal was detected by only one or two markers. Results showed that the proportion of PCR MRD-positive results obtained with only one marker was 21% in the FCM MRD-negative group compared to 3% in the FCM MRD-positive group.

The discrepancy was greatest (i.e. one result was <0.01% and the other >0.1%) in 35 out of 1115 cases (3.1%) at different time points (n=5 at day 15, n=26 at day 33, and n=4 at day 78). However, we did not find any specific feature, i.e. gender, age, white blood cell counts, morphological remission at day 33, DNA index, t(4;11), or final risk stratification.

The positive predictive value of FCM relative to PCR results was high at each time point, being 98%, 74% and 77% at days 15, 33 and 78, respectively. The negative predictive value of FCM was 22%, 69% and 87% at days 15, 33 and 78, respectively.

Triple-comparisons (polymerase chain reaction versus flow cytometry on mononucleated cells or total nucleated cells)

In 266 samples, the overall concordance between FCM and PCR was high (91.3% with FCM^{NC}; 89% with FCM^{MNC}) at the cut-off of 0.01% with most divergent samples being FCM-negative and PCR-positive (Table 2).

The proportions of MRD-positive samples decreased with time during the follow-up because of declining levels of MRD load. Consequently, the sensitivity of FCM declined substantially from day 15 to day 78 (Table 2). In samples that were positive for MRD according to both methods, the proportions measured by PCR were 3.4 times higher than those obtained by FCM^{NC} (SD 4.9; n=73), and 2.9 times higher than those of FCM^{MNC} (SD 5.8; n=75) (Figure 2).

The specificity of FCM, as assessed in samples with neg-

Table 1. Concordance in MRD detection and performance of FCM as compared to PCR at different time points in 2701 bone marrow samples.

	PCR-MRD Day 15 (n. of samples)			PCR-MRD Day 33 (n. of samples)			PCR-MRD Day 78 (n. of samples)		
	PCR ≥0.01%	PCR <0.01%	Total	PCR ≥0.01%	PCR <0.01%	Total	PCR ≥0.01%	PCR <0.01%	Total
FCM-MRD									
FCM ≥0.01%	388	6	394	223	80	303	48	14	62
FCM <0.01%	60	17	77	248	564	812	133	920	1053
Total	448	23	471	471	644	1115	181	934	1115
FCM sensitivity	388/448 = 87%			223/471 = 47%			48/181 = 27%		
FCM specificity	17/23 = 74%			564/644 = 88%			920/934 = 99%		
Concordance rate	405/471 = 86%			787/1115 = 70%			968/1115 = 87%		
Overall concordance rate				2160/2701 = 80%					

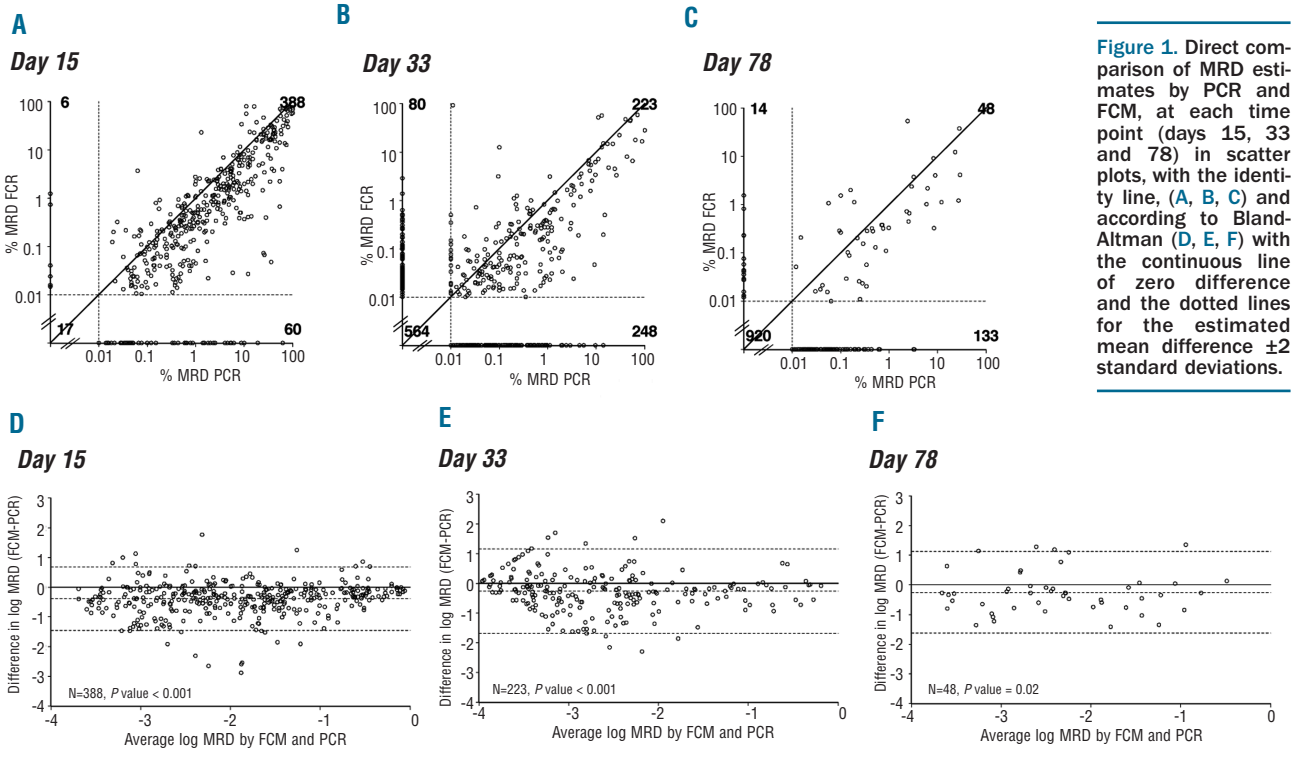


Figure 1. Direct comparison of MRD estimates by PCR and FCM, at each time point (days 15, 33 and 78) in scatter plots, with the identity line, (A, B, C) and according to Bland-Altman (D, E, F) with the continuous line of zero difference and the dotted lines for the estimated mean difference ±2 standard deviations.

ative PCR MRD, was high overall (FCM^{NC} 97.2%, FCM^{MNC} 92.7%, Table 2). Notably, the specificity of both FCM-approaches at day 78 was around 99%.

The overall correlation of MRD estimates by PCR and by FCM with both cell preparations is shown in Figure 2. The concordance between the two FCM assays was 96% (255/266) in positive-negative correlations, and 91% (242/266; Figure 2) using the cut-off of 0.01%. Of 24 divergent samples at the 0.01% cut-off, 11 samples were negative by FCM^{NC} but positive by FCM^{MNC}, mostly at very low levels of MRD. When limiting the FCM^{MNC} assessment to only 3×10^5 cells, as for the FCM^{NC} assay, seven of these 11 samples were MRD negative. Hence, most of the increase in sensitivity was related to the number of cells acquired. The MRD levels measured by FCM^{MNC} had a 1.85 times higher mean than those obtained by FCM^{NC} (SD 1.86; among 86 paired positive samples).

Clinical outcome of patients with concordant or discordant results at day 33

The concordance rate between FCM and PCR MRD data was lowest on day 33. This prompted an outcome analysis in four subgroups of patients with concordant or discordant results by FCM and PCR (Figure 3). Using the cut-off of 0.01%, a very good outcome (5-year EFS 91.6%) was found in the cases concordantly negative for MRD (n=564), and a poor outcome (5-year EFS 50.9%) in those concordantly positive for MRD (n=223). Interestingly, the majority of discordant cases (248 discordant FCM negative and PCR positive patients) showed a better outcome (5-year EFS 77.1%) than concordant FCM positive and PCR positive cases (5-year EFS 50.9%; $P<0.001$). This better outcome is only partially explained by the fact that discordant cases tended to have a low level of MRD by PCR (*Online Supplementary Figure S1*). In fact, 180 out of 248 patients had PCR-MRD values <0.1% and yet the remaining 68 patients with higher values of PCR-MRD had a relatively good outcome (Figure 3, continuous lines).

Among negative PCR MRD cases, discordant positive FCM MRD values identified a subgroup of 80 patients with relatively poor outcome (5-year EFS 81.9%) as compared to concordant cases which had FCM MRD negative results (5-year EFS 91.6%, $P=0.02$). The outcome of these cases varied slightly depending on the level of PCR MRD (Figure 4). Very similar results were observed at day 78 (*data not shown*).

Discussion

We report here a comparison of FCM and PCR for MRD detection in a large subset of children with ALL who were consecutively enrolled in the AIEOP-BFM ALL 2000 multicenter study. Importantly, the concordance rate between these two techniques was assessed prospectively according to their routine and independent application.

All previous studies comparing FCM and PCR used a 0.01% threshold to define MRD positivity.[26-30] By applying this criterion to our series, the overall concordance rate was 80%, with the lowest concordance rate at day 33 (70%). Our results concur with those reported by Malec *et al.*,[28] but contradict others of studies indicating a concordance higher than 90%.[26,27,29,30] Several differences between these studies and ours may explain this. Kerst *et al.*[29] and Ryan *et al.*[30] considered many fewer paired samples than we did in our multicenter study. Furthermore, those authors used MNC as the starting material for both FCM and PCR, while, for the

Table 2. Triple concordance between FCM^{NC}, FCM^{MNC} and PCR in MRD detection at different time points in 266 bone marrow samples.

	PCR-MRD											
	Day 15 (n. of samples)			Day 33 (n. of samples)			Day 52 (n. of samples)			Day 78 (n. of samples)		
	PCR ≥0.01%	PCR <0.01%	Total	PCR ≥0.01%	PCR <0.01%	Total	PCR ≥0.01%	PCR <0.01%	Total	PCR ≥0.01%	PCR <0.01%	Total
FCM-MRD NC												
FCM ≥0.01%	46	1	47	17	1	18	3	2	5	3	1	4
FCM <0.01%	4	4	8	9	42	51	4	46	50	1	82	83
Total	50	5	55	26	43	69	7	48	55	4	83	87
FCM sensitivity	46/50 = 92%			17/26 = 65%			3/7 = 43%			3/4 = 75%		
FCM specificity	4/5 = 80%			42/43 = 98%			46/48 = 96%			82/83 = 99%		
Concordance rate	50/55 = 91%			59/69 = 86%			49/55 = 89%			85/87 = 98%		
Overall concordance rate	243/266 = 91.3%											

	PCR ≥0.01%	PCR <0.01%	Total	PCR ≥0.01%	PCR <0.01%	Total	PCR ≥0.01%	PCR <0.01%	Total	PCR ≥0.01%	PCR <0.01%	Total
FCM-MRD MNC												
FCM ≥0.01%	47	1	48	18	5	23	3	6	9	3	1	4
FCM <0.01%	3	4	7	8	38	46	4	42	46	1	82	83
Total	50	5	55	26	43	69	7	48	55	4	83	87
FCM sensitivity	47/50 = 92%			18/26 = 69%			3/7 = 43%			3/4 = 75%		
FCM specificity	4/5 = 80%			38/43 = 88%			42/48 = 88%			82/83 = 99%		
Concordance rate	51/55 = 93%			56/69 = 81%			45/55 = 82%			85/87 = 98%		
Overall concordance rate	237/266 = 89.0%											

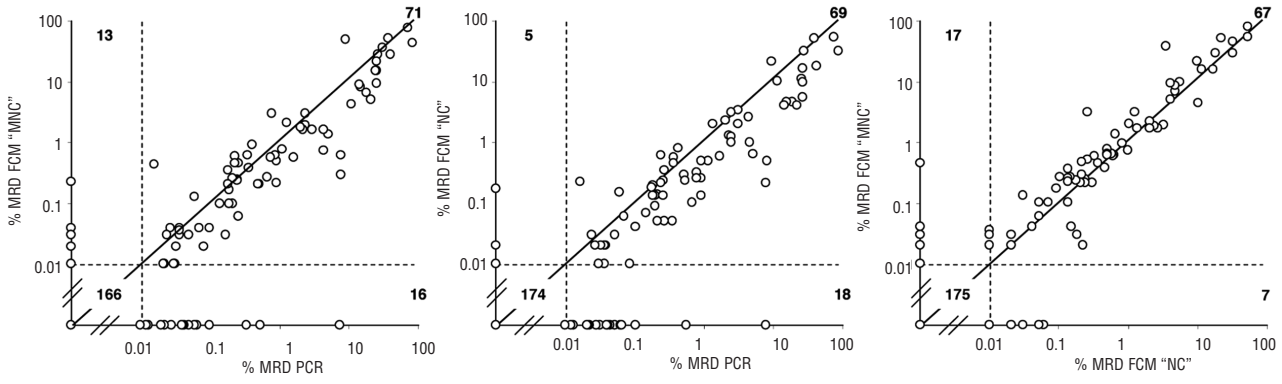


Figure 2. Triple comparison of PCR-MRD (based on MNC) *versus* FCM-MRD (based on either MNC or NC).

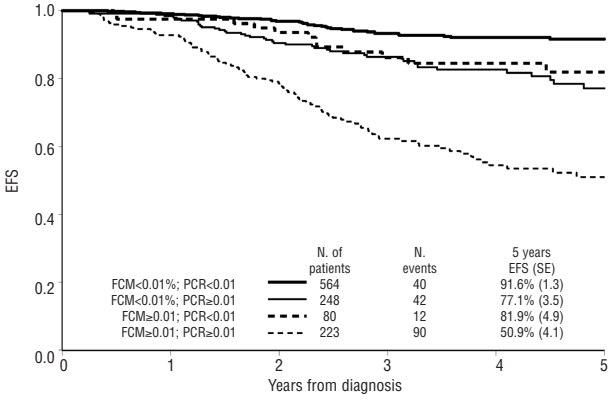


Figure 3. Event-free survival in 1115 patients treated in the AIEOP BFM-ALL 2000 trial, stratified into four groups according to concordant or discordant MRD results obtained on day 33 bone marrow by the simultaneous application of both PCR and FCM.

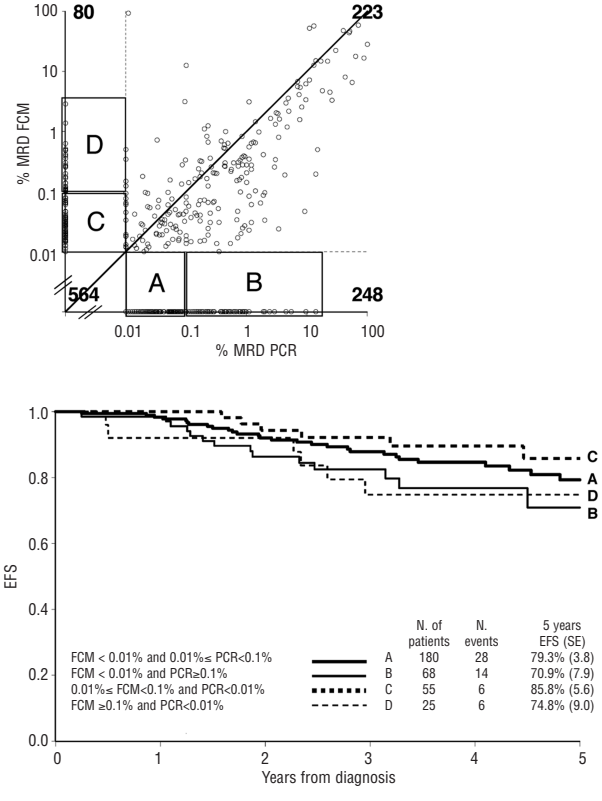


Figure 4. Event-free survival in 328 patients treated in the AIEOP-BFM ALL 2000 trial who had discordant MRD results by PCR and FCM in bone marrow at day 33. Four subgroups of patients with different levels of MRD by either PCR or FCM are represented. For clarity each subgroup can also be identified according to the level of MRD in the scatter plot (see also Figure 1, panel B).

first time, we applied FCM and PCR on total NC and MNC, respectively, using internationally standardized methods.[18,19,26] Of note, divergent results in our study were uniformly distributed among centers; no correlations were found between concordance rate and time delay from collection to analysis; in particular, concordance rates did not change substantially whether the references laboratory processed samples from the local clinic or samples from others clinical centers (overall concordance of 82% and 79%, respectively).

It should also be noted that the overall concordance rate is strongly affected by the number of negative samples. In a study by Neale *et al.*, a concordance rate of 96.7% was obtained in a series including 87.3% (1200 out of 1375) double-negative samples. In our series, we obtained an overall concordance of 80%, but with a lower prevalence of such double-negative samples (1501 out of 2701, 55.6%). Hence, when limiting the analysis to samples with PCR-MRD ≥0.01%, the overall sensitivity of FCM was 60% while its overall specificity was 94%, indicating that the primary difference between FCM and PCR is sensitivity; however, others factors may potentially affect the concordance between FCM and PCR in MRD measurements.[20-25,38]

A few highly discordant results not attributable only to the sensitivity were observed in our study: four samples (three collected on day 15 and one on day 33) had 10-50% MRD by PCR and were negative by FCM; these samples showed bone marrow morphology with a large prevalence of marked apoptosis but without clear evidence of preserved leukemic cells, indicating that PCR amplification of dead cells could have occurred. Those cases were still highly positive at the subsequent time points. By contrast, in three samples (one collected on day 15 and two on day 33) we observed ≥1% MRD by FCM and no MRD by PCR, indicative of a possible evolution of clonal PCR targets, although this could not be formally demonstrated. These patients showed very low MRD values at subsequent time

points. Finally, in one sample collected on day 78, with >1% MRD by FCM and no MRD by PCR, massive marrow regeneration was observed, suggesting a false positive FCM result due to the presence of hematogones with an immunophenotype resembling that of leukemic blasts.

Although in this study similar cell inputs were used in both assays (PCR: DNA of approximately 2.2×10^5 MNC, FCM: 3×10^5 NC) the criteria for assessing MRD positivity were different. While a cluster of at least ten cells with related characteristics had to be detected by FCM to qualify a sample as MRD positive, PCR positivity theoretically relates to a single copy of the target sequence against a background of non-specific amplification.[18] Thus, at the level of 10^{-4}, both methods may be considered equivalent in detecting the amount of MRD (10 leukemic cells in a background of 100000 normal cells). At very low MRD levels, however, it may not always be clear by PCR technology whether the signal observed is due to specific (but not quantifiable) amplification from leukemic cell DNA or from non-specific amplification of normal DNA. Despite this, in protocols aiming at recognizing patients at very low risk, it may be preferable to prevent false-negative MRD results to ensure intensive treatment for all patients. Accordingly, we found that with decreasing levels of MRD, FCM results were more frequently negative in samples qualified positively by PCR. Yet, most discordances were found in samples with low levels of MRD (i.e. <0.1%), so, according to current criteria,[9,31] identification of high risk patients would not be affected.

In previous studies, most authors used MNC as the starting material for both FCM and PCR. We found that the impact of using either MNC or NC on the concordance of FCM with PCR results was minimal, and it was mainly due to the higher amount of input and number of acquired cells. FCM results from MNC and NC were divergent in only 1.5% of samples (4 out of 266), potentially due to different sample preparations or color set-up (i.e. seven- instead of four-color FCM). However, increasing cell acquisition from 3×10^5 to 5×10^5 cells was still not enough to yield a sensitivity similar to that of the PCR technology. Developments in FCM including high-performance data processing and storage equipment may allow us to further increase sensitivity.

During the last 15 years, measuring MRD in the bone marrow of children with ALL has made it possible to identify patients with different prognoses.[4-8,39] Hence, MRD measurement is now widely applied to assign patients to different risk categories in many front-line treatment protocols. Since PCR and FCM showed the most discordant results on day 33, we assessed outcome with combined MRD results derived from both methods at this time point. Interestingly, patients who were MRD-positive on day 33 (i.e. ≥0.01%) by PCR, but negative by FCM, had a better EFS (77.1% at 5 years) than patients with MRD above this threshold with both techniques (50.9% at 5 years; $P<0.001$). On the other hand, patients who were PCR-MRD negative and FCM-MRD positive on day 33 had a 5-year EFS of 81.9% compared to 91.6% of those with MRD below this threshold by both techniques. However, the PCR/FCM discordant cases (PCR positive/FCM negative and *vice versa*) had an outcome much closer to the double negative than to the double positive cases, suggesting that at this time point the presence of very low levels of residual leukemia cannot be strongly predictive of worse outcome.

Although, in general, cut-off levels for technique comparison (i.e. 0.01%) might not necessarily correspond to MRD levels with clinical impact, the intermediate outcome observed (at low MRD levels) in discordant groups may represent the effect of intermediate levels of MRD, at the limit of the sensitivity of both FCM and PCR. These data sustain a potentially complementary role of the two technologies, when applicable and economically sustainable, for further improvement of treatment tailoring.

To our knowledge this is the first report comparing MRD results by FCM and PCR technologies in a multicenter setting with the largest series of patients ever analyzed. We conclude that concordance rates between FCM and PCR largely depend on the time point of the analyses, when varying regenerating cell backgrounds as well as diverse tumor burdens create the conditions for a different capability of the two methods to detect residual leukemic cells. Indeed, most discordances occur at the lowest levels of MRD. If the two technologies should be applied contemporary for clinical decisions, in case of discrepant results the highest (reproducible) value should be assumed, to avoid the risk of under treatment of patients. However, both FCM and PCR methods independently provide strong prognostic information at appropriate time points, and can be used for patient risk stratification, as shown in particular for BFM-based protocols in various studies.[9,39,40] AIEOP-BFM ALL 2000 trial patients were stratified according to PCR-MRD levels on days 33 and 78 of treatment using MRD negativity and high positivity, respectively.[9] For future treatment tailoring, FCM MRD measured in the subsequent protocol on day 15 will contribute to refining stratification, based on recent findings.[31]

The choice of the methods to be used depends on expertise, resources and design of the clinical trial, established in each setting.

Authorship and Disclosures

The information provided by the authors about contributions from persons listed as authors and in acknowledgments is available with the full text of this paper at www.haematologica.org.

Financial and other disclosures provided by the authors using the ICMJE (www.icmje.org) Uniform Format for Disclosure of Competing Interests are also available at www.haematologica.org.

References

1. Brisco MJ, Condon J, Hughes E, Neoh SH, Sykes PJ, Seshadri R, et al. Outcome prediction in childhood acute lymphoblastic leukaemia by molecular quantification of residual disease at the end of induction. Lancet. 1994;343(8891):196-200.
2. van Dongen JJ, Seriu T, Panzer-Grümayer ER, Biondi A, Pongers-Willemse MJ, Corral L, et al. Prognostic value of minimal residual disease in acute lymphoblastic leukaemia in childhood. Lancet. 1998;352(9142):1731-8.
3. Cavé H, van der Werff ten Bosch J, Suciu S, Guidal C, Waterkeyn C, Otten J, et al. Clinical significance of minimal residual disease in childhood acute lymphoblastic leukemia. European Organization for Research and Treatment of Cancer Childhood-Leukemia Cooperative Group. N Engl J Med. 1998;339(9):591-8.
4. Coustan-Smith E, Sancho J, Behm FG, Hancock ML, Razzouk BI, Ribeiro RC, et al. Prognostic importance of measuring early clearance of leukemic cells by flow cytometry in childhood acute lymphoblastic

leukemia. Blood. 2002;100(1):52-8.
5. Dworzak MN, Fröschl G, Printz D, Mann G, Pötschger U, Mühlegger N, et al. Prognostic significance and modalities of flow cytometric minimal residual disease detection in childhood acute lymphoblastic leukemia. Blood. 2002;99(6):1952-8.
6. Borowitz MJ, Devidas M, Hunger SP, Bowman WP, Carroll AJ, Carroll WL, et al. Clinical significance of minimal residual disease in childhood acute lymphoblastic leukemia and its relationship to other prognostic factors: a Children's Oncology Group study. Blood. 2008;111(12):5477-85.
7. Knechtli CJ, Goulden NJ, Hancock JP, Grandage VL, Harris EL, Garland RJ, et al. Minimal residual disease status before allogeneic bone marrow transplantation is an important determinant of successful outcome for children and adolescents with acute lymphoblastic leukemia. Blood. 1998;92(11): 4072-9.
8. Bader P, Hancock J, Kreyenberg H, Goulden NJ, Niethammer D, Oakhill A, et al. Minimal residual disease (MRD) status prior to allogeneic stem cell transplantation is a powerful predictor for post-transplant outcome in children with ALL. Leukemia 2002; 16(9):1668-72.
9. Conter V, Bartram CR, Valsecchi MG, Schrauder A, Panzer-Grümayer R, Möricke A, et al. Molecular response to treatment redefines all prognostic factors in children and adolescents with B-cell precursor acute lymphoblastic leukemia: results in 3184 patients of the AIEOP-BFM ALL 2000 study. Blood. 2010;115(16):3206-14.
10. Eckert C, Biondi A, Seeger K, Cazzaniga G, Hartmann R, Beyermann B, et al. Prognostic value of minimal residual disease in relapsed childhood acute lymphoblastic leukaemia. Lancet. 2001;358(9289):1239-41.
11. Campana D. Determination of minimal residual disease in leukaemia patients. Br J Haematol. 2003;121(6):823-38.
12. Szczepa ski T, Orfão A, van der Velden VH, San Miguel JF, van Dongen JJ. Minimal residual disease in leukaemia patients. Lancet Oncol. 2001;2(7):409-17.
13. Foroni L, Harrison CJ, Hoffbrand AV, Potter MN. Investigation of minimal residual disease in childhood and adult acute lymphoblastic leukaemia by molecular analysis. Br J Haematol. 1999;105(1):7-24.
14. Pongers-Willemse MJ, Seriu T, Stolz F, D'Aniello E, Gameiro P, Pisa P, 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):110-8.
15. Campana D, Coustan-Smith E. Advances in the immunological monitoring of childhood acute lymphoblastic leukaemia. Best Pract Res Clin Haematol. 2002;15(1):1-19.
16. van der Velden VH, Panzer-Grümayer ER, Cazzaniga G, Flohr T, Sutton R, Schrauder A, et al. Optimization of PCR-based minimal residual disease diagnostics for childhood acute lymphoblastic leukemia in a multi-center setting. Leukemia. 2007;21(4):706-13.
17. van der Velden VH, Cazzaniga G, Schrauder, Hancock J, Bader P, Panzer-Grumayer ER, et al. Analysis of minimal residual disease by Ig/TCR gene rearrangements: guidelines for interpretation of real-time quantitative PCR data. Leukemia. 2007;21(4):604-11.
18. Flohr T, Schrauder A, Cazzaniga G, Panzer-Grümayer R, van der Velden V, Fischer S, et al. Minimal residual disease-directed risk stratification using real-time quantitative PCR analysis of immunoglobulin and T-cell receptor gene rearrangements in the international multicenter trial AIEOP-BFM ALL 2000 for childhood acute lymphoblastic leukemia. Leukemia. 2008;22(4):771-82.
19. Dworzak MN, Gaipa G, Ratei R, Veltroni M, Schumich A, Maglia O, 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):331-40.
20. Lúcio P, Parreira A, van den Beemd MW, van Lochem EG, van Wering ER, Baars E, et al. Flow cytometric analysis of normal B cell differentiation: a frame of reference for the detection of minimal residual disease in precursor-B-ALL. Leukemia. 1999;13(3):419-27.
21. Dworzak MN, Fritsch G, Fleischer C, Printz D, Fröschl G, Buchinger P, et al. Multiparameter phenotype mapping of normal and post-chemotherapy B lymphopoiesis in pediatric bone marrow. Leukemia. 1997;11(8):1266-73.
22. van Wering ER, van der Linden-Schrever BE, Szczepa ski T, Willemse MJ, Baars EA, van Wijngaarde-Schmitz HM, et al. Regenerating normal B-cell precursors during and after treatment of acute lymphoblastic leukaemia: implications for monitoring of minimal residual disease. Br J Haematol. 2000;110(1): 139-46.
23. van Wering ER, Beishuizen A, Roeffen ET, van der Linden-Schrever BE, Verhoeven MA, Hählen K, et al. Immunophenotypic changes between diagnosis and relapse in childhood acute lymphoblastic leukemia. Leukemia. 1995;9(9):1523-33.
24. Borowitz MJ, Pullen DJ, Winick N, Martin PL, Bowman WP, Camitta B. Comparison of diagnostic and relapse flow cytometry phenotypes in childhood acute lymphoblastic leukemia: implications for residual disease detection: a report from the children's oncology group. Cytometry B Clin Cytom. 2005;68(1):18-24.
25. Gaipa G, Basso G, Maglia O, Leoni V, Faini A, Cazzaniga G, et al. Drug-induced immunophenotypic modulation in childhood ALL: implications for minimal residual disease detection. Leukemia. 2005;19(1):49-56.
26. Neale GA, Coustan-Smith E, Pan Q, Chen X, Gruhn B, Stow P, et al. Tandem application of flow cytometry and polymerase chain reaction for comprehensive detection of minimal residual disease in childhood acute lymphoblastic leukemia. Leukemia. 1999;13(8): 1221-6.
27. Neale GA, Coustan-Smith E, Stow P, Pan Q, Chen X, Pui CH, et al. Comparative analysis of flow cytometry and polymerase chain reaction for the detection of minimal residual disease in childhood acute lymphoblastic leukemia. Leukemia. 2004;18(5):934-8.
28. Malec M, van der Velden VH, Björklund E, Wijkhuijs JM, Söderhäll S, Mazur J, et al. Analysis of minimal residual disease in childhood acute lymphoblastic leukemia: comparison between RQ-PCR analysis of Ig/TcR gene rearrangements and multicolor flow cytometric immunophenotyping. Leukemia. 2004;18(10):1630-6.
29. Kerst G, Kreyenberg H, Roth C, Well C, Dietz K, Coustan-Smith E, et al. Concurrent detection of minimal residual disease (MRD) in childhood acute lymphoblastic leukaemia by flow cytometry and real-time PCR. Br J Haematol. 2005;128(6):774-82.
30. Ryan J, Quinn F, Meunier A, Boublikova L, Crampe M, Tewari P, et al. Minimal residual disease detection in childhood acute lymphoblastic leukaemia patients at multiple time points reveals high levels of concordance between molecular and immunophenotypic approaches. Br J Haematol. 2009; 144(1):107-15.
31. Basso G, Veltroni M, Valsecchi MG, Dworzak MN, Ratei R, Silvestri D, et al. Risk of relapse of childhood acute lymphoblastic leukemia is predicted by flow cytometric measurement of residual disease on day 15 bone marrow. J Clin Oncol. 2009;27(31): 5168-74.
32. Basso G, Buldini B, De Zen L, Orfao A. New methodologic approaches for immunophenotyping acute leukemias. Haematologica. 2001;86(7):675-92.
33. Veltroni M, De Zen L, Sanzari MC, Maglia O, Dworzak MN, Ratei R, et al. Expression of CD58 in normal, regenerating and leukemic bone marrow B cells: implications for the detection of minimal residual disease in acute lymphocytic leukemia. Haematologica. 2003;88(11):1245-52.
34. Porwit-MacDonald A, Björklund E, Lucio P, van Lochem EG, Mazur J, Parreira A, et al. BIOMED-1 concerted action report: flow cytometric characterization of CD7+ cell subsets in normal bone marrow as a basis for the diagnosis and follow-up of T cell acute lymphoblastic leukemia (T-ALL). Leukemia. 2000;14(5):816-25.
35. Campana D, Coustan-Smith E. Detection of minimal residual disease in acute leukemia by flow cytometry. Cytometry. 1999;38(4): 139-52.
36. Dworzak MN, Fröschl G, Printz D, Zen LD, Gaipa G, Ratei R, et al. CD99 expression in T-lineage ALL: implications for flow cytometric detection of minimal residual disease. Leukemia. 2004;18(4):703-8.
37. Bland, J.M. & Altman, D.G. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet. 1986;1(8476):307-10.
38. Gaipa G, Basso G, Aliprandi S, Migliavacca M, Vallinoto C, Maglia O, et al. Prednisone induces immunophenotypic modulation of CD10 and CD34 in nonapoptotic B-cell precursor acute lymphoblastic leukemia cells. Cytometry B Clin Cytom. 2008;74(3):150-5.
39. 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):790-4.
40. Schrappe M, Valsecchi MG, Bartram CR, Schrauder A, Panzer-Grümayer R, Möricke A, et al. Late MRD response determines relapse risk overall and in subsets of childhood T-cell ALL: results of the AIEOP-BFM-ALL 2000 study. Blood. 2011;118(8):2077-84.